

Open Research Online

The Open University's repository of research publications and other research outputs

Skills for Creativity in Graphic Design: Testing the relationship between visualisation, written comprehension, and graphic design creativity

Thesis

How to cite:

Jeffries, Karl K. (2017). Skills for Creativity in Graphic Design: Testing the relationship between visualisation, written comprehension, and graphic design creativity. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2016 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000c462>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Skills for Creativity in Graphic Design: Testing the relationship
between visualisation, written comprehension,
and graphic design creativity

A thesis presented for the degree of Doctor of Philosophy
at The Open University

By

Karl K. Jeffries

MA, Manchester Metropolitan University

BA (Hons) Glasgow School of Art

Department of Engineering and Innovation

Faculty of Science, Technology, Engineering and Mathematics

The Open University

September, 2016

Abstract

This thesis explores the relationship between skills, creativity and domain. It is situated within an evolving topic of design creativity; an emerging field that interfaces between *creativity research*, which has often occurred in the field of psychology, and *design research* often associated with the fields of engineering, art and design.

Through five interconnected studies, and the domain of graphic design as the basis for experimentation, the research culminates in testing the relationship between creative output and skills considered important for competent performance in graphic design.

How to assess graphic design creativity in a manner credible to both the creativity research and design research community has been a core challenge for this thesis. The method selected - the Consensual Assessment Technique (CAT) - before this thesis had yet to be used as a reliable measure of graphic design creativity in experimental research. For this reason, three of the studies were directed at establishing the reliability of the CAT.

- One undertook the first systematic literature review of the CAT in design research;
- a second tested the assumption that a graphic design CAT would show acceptable levels of inter-rater reliability (as researchers had found in other domains of design);
- a third focused on optimising the CAT to measure graphic design creativity.

Additionally, given the wide range of graphic design competencies available, a conceptual framework was developed to prioritise those most important, and suitable, for experimental study. Two were identified (mental visualisation, and written comprehension) and each was tested for its relationship to graphic design creativity.

Where other researchers have suggested or found a relationship, in this study, no significant correlations could be evidenced for either mental visualisation or written comprehension with creative output. Such a finding is likely counter-intuitive to many in art and design, thus, the implications of this study impact upon debates within design creativity research, design education, and graduate skill gaps for the Creative Industries.

Contents

Abstract	1
Contents	2
List of Figures.....	8
List of Tables	9
Acknowledgments	11
Chapter 1: Introduction.....	12
1.1 The Theoretical Context of this Research.....	13
1.2 The Educational Context of this Research.....	15
1.2.1 Higher Education for the Creative Industries	17
1.3 The Economic Context of this Research.....	21
1.4 Outline of Thesis Structure and Chapters.....	24
Chapter 2: Definitions and Relationships Between Domain, Skills and Creativity.....	27
2.1 Introduction to Chapter 2.....	27
2.2 Defining Domain.....	28
2.2.1 Definitions of Domain	29
2.2.2 Domains of Mind	29
2.2.3 Occupational Roles.....	29
2.2.4 Microdomains.....	30
2.2.5 Which Definition to Use?.....	30
2.2.6 Related Occupations.....	30
2.2.7 Single Occupations	31
2.2.8 Domain Identity of Participants: Designer or Artist	31
2.2.9 Definition of Domain in this Study	32
2.3 Defining Skills.....	34
2.3.1 Conflated Usage	34
2.3.2 Talent and Giftedness	36
2.3.3 Expertise	36
2.3.4 Competency	37
2.3.5 Abilities	37
2.3.6 Skills.....	38
2.3.7 Tasks.....	38
2.3.8 Knowledge	39
2.3.9 Definition of Skills in this Study	40
2.4 Define Creativity	42
2.4.1 Eastern and Western Definition of Creativity	42
2.4.2 The 4 P's of Creativity Research: Product, Process, Person and Press	43

2.4.3 Historical and Personal Creativity	43
2.4.4 Creative Output	44
2.4.5 Protocol Studies	45
2.4.6 Self-Report Measures	45
2.4.7 Divergent Thinking Tests and the Torrance Tests of Creative Thinking	45
2.4.8 Creativity Assessment by Domain Experts, and the Consensual Assessment Technique	46
2.4.9 Definition of Creativity used in this Study	49
2.5 Key Relationship: Creative and Domain	50
2.5.1 Domain-General Theory	50
2.5.2 Domain-Specific Theory	50
2.5.3 Hybrid Models	51
2.5.4 Conclusion	51
2.6 Key Relationship: Domain and Skill	52
2.6.1 O*NET Graphic Design	53
2.6.2 National Occupational Standards.	55
2.6.3 BTEC Specialist	57
2.6.4 BTEC Higher Nationals (Graphic Design) Level 4 & 5	58
2.6.5 NASAD	60
2.6.6 Conclusion	61
2.7 Key Relationship: Skill and Creativity	61
2.7.1 A Linear Negative Correlation	62
2.7.2 A Linear Positive Correlation	62
2.7.3 A Non-Linear Positive Correlation	63
2.7.4 A Non-Linear Negative Correlation	64
2.7.5 Comparable Shapes of Correlations	65
2.7.6 Influence of Domain on Skills/Creativity Relationship	65
2.8 Core Relationship: Skills, Creativity and Domain	66
2.8.1 Skills for Creativity in Graphic Design	68
Chapter 3, Study 1: A Systematic Review of the CAT in Design Research	72
3.1 Introduction to Study 1	72
3.2 Systematic Review	73
3.3 Method	74
3.3.1 Why use Google Scholar Rather than Other Databases?	74
3.3.2 Downloading Multiple Citations from Google Scholar	75
3.3.3 Cleaning the Reference Manager Database	75
3.3.4 PsycINFO	76
3.3.5 Web of Knowledge	76

3.3.6 Design Research Journals	77
3.4 Results	77
3.4.1 CAT Citations	77
3.4.2 CAT Citations Within Design Journals	77
3.4.3 Key Journals Publishing CAT Studies	79
3.4.4 Key Authors Publishing CAT Studies.....	79
3.4.5 CAT Studies Related to Graphic Design.....	80
3.5 Discussion	80
3.5.1 Creative Process Versus Creative Product.....	81
3.5.2 Method and Methodological Concerns with a Social Consensus View of Creativity	83
3.5.3 Criterion Versus Normative Assessment	86
3.5.4 Use of CAT in Graphic Design	90
3.6 Conclusion	92
Chapter 4, Study 2: A Conceptual Framework for Prioritising Graphic Design Competencies.	94
4.1 Introduction to Study 2	94
4.2 Literature Specific to this Chapter.....	94
4.2.1 CHC Theory	96
4.3 Method	106
4.4 Results	109
4.4.1 Mapping ONET Graphic Design Abilities to CHC Theory.....	109
4.4.2 ONET/CHC Cross References with Graphic Design Competencies	111
4.4.3 Overall Ranking and Priorities.	114
4.5 Discussion	115
4.5.1 Operational Measures for Written Comprehension and Visualisation	117
4.5.2 Operational Measures for Written Comprehension	118
4.5.3 Operational Measures for Visualisation	122
4.5.4 Limitations of These Findings.	130
4.6 Conclusions.....	130
Chapter 5, Study 3: Is the CAT a Reliable Method of Assessing Graphic Design Creativity?	132
5.1 Introduction to Study 3	132
5.1.1 A Task for a Graphic Design CAT.....	132
5.2 Watching Words Move Task	135
5.2.1. Method	136
5.2.2 Results	142
5.2.3 Discussion of WWM Task Results	144
5.3 T-shirt Design Task	146
5.3.1 Method	147
5.3.2 Results	151

5.3.3 Discussion T-shirt Task Results.....	152
5.4 Broad Discussion	154
5.4.1 Technical Execution	155
5.4.2 Suitable Judges.....	156
5.4.3 Range of Artwork	157
5.4.4 T-shirt Task	158
5.5 Conclusions.....	160
Chapter 6, Study 4: Optimising the CAT to Measure Graphic Design Creativity.....	162
6.1 Introduction to Study 4	162
6.2 Literature Specific to this Chapter.....	162
6.3 Method	164
6.3.1 Broad Research Design	164
6.3.2 Ethical Considerations	166
6.3.3 Tasks.....	168
6.3.4 Instruction to Judges	169
6.3.5 Judges.....	181
6.3.6 Procedures.....	184
6.4 Results	185
6.4.1 Type Task	185
6.4.2 Image Task	186
6.5 Discussion.....	188
6.5.1 Diversity of Creativity Within a CAT Grouping	189
6.5.2 Technical Ratings.....	191
6.5.3 Rankings, not Just Alpha	191
6.5.4 To Include a Caveat or not?.....	192
6.6 Conclusion	196
Chapter 7, Study 5: The Relationship Between Graphic Design Competencies and Creative Output	198
7.1 Introduction to Study 5	198
7.2 Background to this Chapter	199
7.2.1 A Range of Graphic Design Experience	200
7.2.2 Graphic Design Creativity and Novice Participants	201
7.2.3 Visualisation and Written Comprehension	208
7.2.4 The Influence of General Intelligence	209
7.2.5 An Appropriate Size of Group	210
7.2.6 Main Research Question and Subordinates	211
7.3 Method	212
7.3.1 Broad Research Design	212

7.3.2 Identification of Experts, Novices and Intermediates	213
7.3.3 Participants	215
7.3.4 Measures.....	217
7.3.5 Judges.....	220
7.3.6 Ethics, Data Protection, Location and Recompense	224
7.4 Results	226
7.4.1 Inter-Rater Reliability Scores for Image and Type Tasks Creativity	227
7.4.2 Normality of Distribution	228
7.4.3 Correlations Between Creativity, Aesthetic Appeal and Technical Execution	232
7.4.4 Correlations Between Type and Image Creativity	236
7.4.5 Correlations Between Creativity and Visualisation	237
7.4.6 Correlations Between Creativity and Written Comprehension.....	241
7.4.7 Correlation Between Creativity and General Intelligence	246
7.5 Discussion	248
7.5.1 Graphic Design Creativity and Visualisation	249
7.5.2 Graphic Design Creativity and Written Comprehension	250
7.5.3 A Graphic Design CAT Without Pre-Selection of Artwork.....	253
7.5.4 A Combined Creativity Score for Type and Image.....	254
7.5.5 Correlations Between Creativity, Technical Execution and Aesthetic Appeal	256
7.5.6 Emergent Considerations.....	258
7.6 Conclusion	260
Chapter 8: General Discussion and Conclusion	262
8.1 Introduction	262
8.2 The Art and Science of Creativity Research	262
8.3 CHC Theory and Graphic Design Competencies	265
8.4 Growth of CAT Update.....	266
8.5 Novelty of CAT to Design Studies.....	268
8.6 Creativity Only Instruction	269
8.7 Correlations Between Creativity, Technical Executions and Aesthetic Appeal	270
8.8 Distinction Between Aesthetic and Technical in Other Design Domains.....	271
8.9 Not Meeting Assumptions	272
8.10 Is Creativity Domain Specific?	272
8.11 Ethics of Showing Creative Outputs	275
8.12 Limitations	280
8.13 Future Studies.....	281
8.14 Conclusion	283
References.....	286
Appendix A: Example of Consent Forms for Study 4	295

Contents

Appendix B: Randomisation of Type Task Artwork for Study 4	296
Appendix C: Randomisation of Image Task Artwork for Study 4.....	297
Appendix D: Image Task Technical Instructions for Study 5	298
Appendix E: Example of Information Sheet for Study 5.....	299

List of Figures

Figure 1: Key relationships between skill, creativity and domain.....	27
Figure 2: A negative linear relationships for skills/knowledge relative to creativity	62
Figure 3: A positive linear relationships for skills/knowledge relative to creativity	63
Figure 4: A positive non-linear relationships for skills/knowledge relative to creativity.....	64
Figure 5: A non-linear negative relationships for skills/knowledge relative to creativity	65
Figure 6: Number of CAT citation (per year) in design journals relative to other journals	77
Figure 7: Images of artworks selected for Dineen and Niu (2008)	91
Figure 8: O*NET's prioritisation of graphic design abilities.....	95
Figure 9: (Left) Form Board Test, (Right) Revised Minnesota Paper Form Board Test	124
Figure 10: The Paper Folding Test (VZ-2)	124
Figure 11: (top) PSVT:R, (Bottom) Vandenberg & Kuse's Mental Rotation Test	126
Figure 12: Example from Watching Words Move	136
Figure 13: Artwork selected for the WWM task.....	138
Figure 14: Artwork selected for the WWM task (continued)	139
Figure 15: Comparison of expert scores with novice/intermediate scores for the WWM task	143
Figure 16: Example of restricted rating sheet	145
Figure 17: A PowerPoint drag and drop rating sheet was used for the T-shirt Task.....	151
Figure 18: Two examples of poster design used	163
Figure 19: Six examples of artwork (none of these examples were part of this study).	169
Figure 20: Option 1	175
Figure 21: Option 2	176
Figure 22: Option 3	176
Figure 23: Rating sheet developed to graphically reinforce Kaufman's et al. CAT protocol	185
Figure 24: Scatterplot for Type Task.....	186
Figure 25: Scatterplot for Image Task.....	187
Figure 26: Alpha and r in relation to task and instructions	188
Figure 27: Information on Technical Support for the Image Task.....	205
Figure 28: Information on Technical Support for the Type Task.....	206
Figure 29: Screen Shot of G-Power analysis for a one-tailed correlational design	211
Figure 30: Request for information of level of experience in graphic design	215
Figure 31: Histogram of Nelson-Denny Reading Rate scores.....	231
Figure 32: Histogram of KBIT Non-Verbal scores.....	232
Figure 33: Scree plot for Type Task.....	235
Figure 34: Scree Plot for Image Task	236
Figure 35: Scatter plot of Type Task in relation to ETS Advance Vocabulary Test.....	244
Figure 36: Scatter plot of Type Task to Image Task creativity scores	256
Figure 37: Number of CAT citations updated to 2015	267

List of Tables

Table 1: O*NET's occupation-specific tasks for graphic design	53
Table 2: O*NET's abilities for graphic design	54
Table 3: O*NET's knowledge for graphic design	54
Table 4: O*NET's skills for graphic design	54
Table 5: CCS's Create Visual Designs performance criteria	56
Table 6: CCS's Create Visual Designs knowledge required	57
Table 7: BTEC Specialist, Graphic Design unit options	57
Table 8: BTEC Higher Nationals (Graphic Design) Core Units	58
Table 9: BTEC Higher Nationals (Graphic Design) learning outcomes	59
Table 10: NASAD essential competencies for graphic design	60
Table 11: CHC theory	98
Table 12: BTEC Specialist units not included in study 2	108
Table 13: BTEC Specialist units, and learning outcomes, included in study 2	108
Table 14: Mapping of CHC to O*NET	110
Table 15: O*NET Core Tasks cross-referenced to abilities	112
Table 16: NASAD essential competencies cross referenced to abilities	112
Table 17: BTEC Specialist cross-referenced to abilities	113
Table 18: BTEC HN's cross-referenced to abilities	114
Table 19: Comparison of abilities across each model	115
Table 20: Wording of a CAT task given to participants	140
Table 21: Wording for WWM task given to participants	141
Table 22: Instructions give to judges	141
Table 23: Rating sheet provided to judges	142
Table 24: T-shirt instructions given to judges	149
Table 25: Instructions for judges	169
Table 26: Comparison of instructions with Kaufman et al.	170
Table 27: Comparison of instructions with Baer	171
Table 28: Counterbalanced tasks and instructions	182
Table 29: Randomisation of judges into groups	183
Table 30: Collation of judges rating for alpha calculations	183
Table 31: Random allocation of judges into counterbalanced conditions	224
Table 32: Descriptive statistics for calculating Z-skew of Image Task	228
Table 33: Descriptive statistics for calculating Z-skew of Type Task	229
Table 34: Descriptive statistics for calculating Z-skew for visualisation measures	229
Table 35: Descriptive statistics for calculating Z-skew of written comprehension measures	230
Table 36: Descriptive statistics for calculating Z-skew of general intelligence scores	232
Table 37: Guidance on the strength of relationship associated with Pearson's r.	232
Table 38: Correlations for the Image Task	233
Table 39: Correlations for the Type Task	234
Table 40: PCA of Type Task	234
Table 41: Type Task PCA component matrix	235

Table 42: PCA of Image Task	236
Table 43: Image Task PCA component matrix	236
Table 44: Correlations between total creativity and visualisation measures.....	237
Table 45: Correlations controlled for general intelligence	238
Table 46: Correlations between Type Task and visualisation measures	239
Table 47: Correlations controlled for general intelligence	239
Table 48: Correlations between Image Task and visualisation measures	240
Table 49: Correlations controlled for general intelligence	241
Table 50: Correlations between total creativity and written comprehension measures.....	242
Table 51: Correlations controlled for general intelligence	242
Table 52: Correlations between Type Task and written comprehension measures	243
Table 53: Correlations controlled for general intelligence	245
Table 54: Correlations between Image Task and written comprehension measures	245
Table 55: Correlations controlled for general intelligence	246
Table 56: Total creativity correlations for KBIT.....	247
Table 57: Type Task correlations for KBIT.....	247
Table 58: Image Task correlation for KBIT	248
Table 59: Increase in alpha across study 4 and 5.....	270
Table 60: Extract from fieldwork notes	278

Acknowledgments

This research has been supported by several academics at the Open University and the University of Central Lancashire, family and friends. Specifically, I would like to praise both of my supervisors, Dr Alison Green and Dr Theodore Zamenopoulous, who have been a constant source of encouragement, insight, and offered me everything I could have hoped for through a PhD, and so much more. At UCLan, my thanks to Andrew Bainbridge for access to his classes, and discussion of creative tasks for graphic design (both of which were essential to the development of this research). Prof. Lubaina Himid, for supporting all those practical needs (like funding for resources, and attendance at international conferences) and the many intangible aspects of research mentorship that have helped me grow. I am also indebted to Prof. Linden Ball, whose support for my research has been a constant source of encouragement. My thanks to all the participants of each study; for reasons of anonymity you remain unknown here, but you know who you are, and you have my heartfelt thanks; without your support, none of this research would have been possible. Lastly, all of my love to my wife Sue Jeffries, I would not be the person I am today, let alone have been capable of this research, without her by my side.

Chapter 1: Introduction

This thesis is about exploring further the relationship between skills, creativity and domain. It is about identifying those skills that a domain considers crucial for competent performance, and testing their relationship (if any) to creative output within a domain. The domain in question is graphic design. The aim is to test the core research question: *To what extent are domain competencies correlated with creative output in graphic design?*

Such a purpose belies complexity and detail, but above this level of scrutiny are broader themes that establish why this question is important. For example, each generation of creativity researchers explores the nature of creativity. In recent decades a key theme has been creativity as a domain-specific or domain-general ability; this thesis contributes further evidence to this theoretical debate through empirical study. Additionally, the interaction between higher education and the development of professional competencies is a topical discussion within academia, industry and government. Creativity, for example, has been described as a core competency for industry (Berland, 2012), or in the context of design and design professions “a key aspect” (Cross, 2006, p12). For other researchers, skills that underpin mental visualisation are important to professional competency and “a crucial part of creative problem solving in the design process” (Bennett, 2015, p321). In relation to competencies and creativity, likewise, this thesis contributes new evidence through empirical study.

In this chapter, to introduce these issues further, the theoretical, educational, and economic context are presented, followed by outlining the overall thesis structure and a summary of each chapter.

1.1 The Theoretical Context of this Research

This thesis is situated within an evolving topic of design creativity. It builds on an emerging field that interfaces between *creativity research*, which has often occurred in the field of psychology, and *design research* often associated with the fields of engineering, art and design.

The development of the American Psychological Association's (APA) Division 10 (Society for the Psychology of Aesthetics, Creativity, and the Arts), for example, is of particular interest. Division 10 was specifically designed as an interdisciplinary section of the APA, focusing on the nexus between psychology and the creative arts. As an area of research, this has seen considerable academic growth in recent years. The Division's journal, *Psychology of Aesthetics, Creativity, and the Arts (PACA)*, had substantial improvement in its impact factor rating in 2014 (3.054, an increase from 1.787). Its standing within general psychology/experimental journals ranked 17 out of 85 journals (up from 41) (P. P. L. Tinio, personal communication to Division 10 members, June 19, 2015) highlighting the increased attention and quality of scholarship that this area of research has attracted.

Equally, much like the history of APA's Division 10, many design researchers and eminent design journals have had a longstanding interest in creativity. The Creativity and Cognition Conference series founded by Ernest Edmonds and Linda Candy in 1993; John Gero's organisation of the Computational and Cognitive Models of Creative Design Conferences from 1989 onwards; the Design Society's first International Conference on Design Creativity (ICDC) in Japan during 2010, with subsequent conferences in the UK, 2012, India, 2015, and the United States, 2016, are all examples.

One question frequently explored has been that of whether or not domain influences creativity (Barbot, Besançon & Lubart, 2016; Julmi & Scherm, 2015; Plucker & Beghetto,

2004). While this type of enquiry could appear purely theoretical, the implications have potential to be significant for creativity training and education.

For example, Baer's 1991 study highlighted that creativity training that increased creativity in poetry did not enable participants to transfer this to short story writing (Baer, 1991, 1993); yet both of these areas would be categorised more broadly under creative writing. The evidence suggested that insights from creativity training did not transfer between domains, even for domains one might consider to be closely associated with each other. The inference was that if creativity is notably influenced by domain knowledge, this could explain why some training programs are more effective than others (Scott, Leritz, & Mumford, 2004); it might also see a theoretical basis for creativity training that exclusively focuses on generic content undermined (Baer, 1998). Such statements were clearly problematic due to much creativity training, historically, been based on domain free, generic content. Moreover, as will be discussed in section 2.3 (Defining Skills) the issue of what precisely constitutes a skill can be confused; especially when researchers and commentators blend their definition of skills with other constructs like knowledge, abilities, talents, and so forth.

Given this, theoretical underpinnings are not without controversy (Plucker, 1998; Han & Marvin, 2002; Plucker et al., 2004; Silvia, Kaufman, & Pretz, 2009; Reiter-Palmon, Illies, Cross, Buboltz, & Nimps, 2009, Baer, 2012) and, at the extremes, two schools of thought have evolved regarding the extent scholars consider creativity to be either domain-specific or domain-general (Han, 2003). In chapter 2, section 2.2, both schools of thought will be described in detail, in addition to the perspective that creativity requires both domain-specific and domain-general competencies (such as divergent thinking) at the same time.

This thesis, focused on domain, skills and creativity, partly stems from the debate on the domain-specificity or generality of creativity. Ivcevic (2009) made the case that next generation creativity theory needed to be established on a better understanding of domain-specificity, how it differs from domain-generality, and that more detailed research was necessary to achieve this. By focused empirical study on the relationship between graphic design creativity and graphic design competencies, the intention is to add to this debate, and the understanding of domain specificity.

Whilst one aspect of academia has been concerned with theoretical implications, debate over the domain-specificity versus domain-generality of creativity also mirrors a dichotomy seen in tensions between the needs of academia and the needs of industry; between those who advocate the need for graduates with more domain-specific skills versus those who highlight the importance of a broader set of conceptual abilities and knowledge. This tension is none the more present than in the example of higher education for the Creative Industries.

1.2 The Educational Context of this Research

The "Creative Industries", a term that began in the 1990's (Bendassolli, P. F., Wood, Kirschbaum & Cunha, 2009), and one closely aligned to the politics of New Labour (Guile, 2010) in the United Kingdom (UK), became a constant theme for government policy in the UK from 1997 (Crossick, 2006), with many initiatives undertaken to support this sector's growth. For example, the Design Skills Advisory Panel's 2007 report *Higher Skills for Higher Value* for the Design Council and Creative & Cultural Skills (Design Skills Advisory Panel, 2007), or *Creativity and Culture: The next ten years* from the Department of Culture, Media and Sport in 2001 (DCMS, 2001b).

While the precise definition is contested, creativity is one core characteristic described as an essential dimension of the Creative Industries, and central to its definition. In 2001 (based on earlier research) the Department for Culture, Media and Sport (DCMS) defined the Creative Industries as, “industries which have their origin in individual creativity, skill and talent and which have a potential for wealth and job creation through the generation and exploitation of intellectual property” (DCMS, 2001a, p.5).

Under this definition they identify 13 sectors to be included: “advertising, architecture, the art and antiques market, crafts, design, designer fashion, film and video, interactive leisure software, music, the performing arts, publishing, software and computer services, television and radio” (DCMS, 2001a, p.5).

Which sectors to include, and which to exclude, was acknowledged as a challenge from the outset, and indeed, eminent members of the scientific community, such as Harry Kroto, found the exclusion of fields like nanotechnology from being considered a creative industry to be conceptually flawed. Without negating the validity of this criticism, the DCMS mapping of the creative industries enabled a strategic focus to be placed on statistical data collection which was used to inform government policy.

In the intervening years from 1998, the DCMS has engaged stakeholders (from the Office for National Statistics (ONS) to the National Endowment for Science, Technology and the Arts (Nesta), and the sector skills councils, Creative Skillset and Creative and Cultural Skills), with the aim to revise the definition and how statistics are collected. Through doing so, the current DCMS definition of the Creative Industries has moved towards occupations considered to be creative. The proportion of those occupations within an industry (which they describe as creative intensity) above a specified threshold (30%) being the basis for

inclusion as a part of the Creative Industries (DCMS, 2016a). Currently, the Creative Industries are as follows:

- Advertising and marketing
- Architecture
- Crafts
- Design: product, graphic and fashion design
- Film, TV, video, radio and photography
- IT, software and computer services
- Publishing
- Museums, galleries and libraries
- Music, performing and visual arts

1.2.1 Higher Education for the Creative Industries

Much of the education required to enter the Creative Industries in the UK has occurred via higher education. Around 73% of those employed in 2008 had achieved a minimum of an undergraduate degree qualification (Skillset, 2009), which was considerably more than the UK national average, and for specific sectors like computer games, and animation, this figure rose to 80%, and 92% respectively. Furthermore, this data suggests that over half of the undergraduate and postgraduate qualifications gained were in subjects related to the specific sector in which graduates found work. In this regard, graduates with qualifications closely aligned to their job roles form a considerable part of the Creative Industries.

However, the suitability of these graduates for employment within the Creative Industries has been questioned and, to this end, the type of content covered by their academic qualifications suggested as a cause for concern. Perhaps the clearest example of this within the Creative Industries can be seen in computer games.

For example, Creative Skillset (previously known as Skillset: The Sector Skills Council for the Audio and Visual industries) followed this theme within their undergraduate course accreditation guidelines for computer games. They highlighted that the increased number

of higher education games related courses in the UK had resulted in "concern being voiced by employers in the computer games industry that graduates from these courses do not have the requisite skills to sufficient level to meet the needs of the industry" (*Skillset, 2008, p.3*), and that accreditation of provision is seen as the best means to rectify this situation. David Braben, the campaign spokesman for Games Up? (a campaign group, sponsored by some of the UK's largest games development studios and trade bodies, to raise the profile of the games industry in Parliament and the media), was quoted as saying "...95% of video gaming degrees are simply not fit for purpose. Without some sort of common standard, like Skillset accreditation, these degrees are a waste of time for all concerned", (Lipsett, 2008, ¶ 14).

To date, such concerns remain, have grown, and are not specific to computer games or the Creative Industries. A 2015 skills survey by the Institution for Engineering and Technology (IET) found that:

Some 53% employers say that a typical new recruit does not meet their 'reasonable expectations' – and year-on-year data shows that employers are growing more and more concerned that recruits do not meet their expectations. Nearly two-thirds report that graduates are the biggest challenge and a similar percentage is concerned that the education system will struggle to keep up with the skills required for technological change. (p. 4).

In addition, HM Treasury's July 2015 report *Fixing the Foundations* predicted that the UK will fall into 28th place out of 33 OECD countries by 2020 (The Organization for Economic Co-operation and Development), behind countries such as France, Germany, United States, Canada and Australia, if it continues to "performs poorly on intermediate professional and technical skills" (p23). Part of the solution suggested has remained the

same for some time: the need to put "employers in the driving seat" (p.23) with regard to technical and professional qualifications. Echoing the views of Dinah Caine (at the time, Chief Executive of Skillset, now Chair of Creative Skillset's Board of Directors) from over a decade earlier (2004a).

In contrast, for academics within Art, Design and Media, greater emphasis being placed on industry's opinion of curriculum content (for example, through Sector Skills Council advisory groups) was a significant concern (Wall et al., 2006). For some, industry led direction, was an attempt to standardise curriculums in Higher Education, and symptomatic of a "mindset" (p.139) that believes UK Art, Design and Media education to be defective, and unable to meet the needs of employers (Macdonald, 2006).

Professor Geoffrey Crossick, (Warden of Goldsmiths, University of London), in his speech to the Royal Society of Arts, highlighted that it was "...important not to assume that employers automatically know best what education their future employees need" (p.12), and that a university education should be about "...developing people not just with the skills to meet today's needs but also the conceptual abilities and imagination to take risks that will generate what is needed in the future" (Crossick, 2006, p.12).

For other commentators, the debate had been less about who controls the curriculum and standards, and more about industry engaging with higher education. The Council for Industry and Higher Education (CIHE) for example published 'The Fuse: Igniting High Growth for Creative, Digital and Information Technology Industries (CDIT) in the UK' (CIHE, 2010). Through the report, it raised a number of the issues around business:

....playing its part in developing graduates capable of leading the UK's CDIT

industries....(for) employers to collaborate closely with the universities that supply them...(and) that industry bodies....should promote volunteer schemes through which professionals can work with students and help them develop the employability skills they need for the CDIT jobs market. (Colman Getty, 2010, p.3).

What each of the different examples above suggest, as a generalization, are contrasting emphases on the relationship of creativity and skills with domain competency: where higher education has focused on creativity as a means to deal with an unpredictable future, industry has focused on the identifiable skill needs of today in order to maintain a competitive advantage now.

Yet, it is unclear whether taking a domain-specific, domain-general, or hybrid approach is most useful. Indeed, the acquisition of domain-specific "skills" may well be the best mechanism to enhance creativity. Equally, focusing on domain-general "conceptual abilities" could be the more fruitful. Alternatively, exploring a combination of these may be the better direction to pursue. From a theoretical and an empirical perspective, organisations and individuals do not know, though strong opinions and convictions (as suggested above) are expressed on either side of this debate.

This thesis does not directly address how creativity can be taught, but it does aim to test (via empirical study) a number of perspectives which underpin the debates above: perspectives about the relationship between creativity and skills; perspectives that shape (knowingly or not) the type of pedagogy and curriculum within design higher education.

1.3 The Economic Context of this Research

Creativity has often been described, and prescribed, as a: means to economic growth; core attribute for employment; key goal of higher education (HM Treasury, 2005; DCMS, 2001b; Universities UK, 2010).

Though inherently linked, the value of theoretical perspectives on creativity and the practicalities of educating for creativity can be a means to an end, the goal being economic. By its nature, such a perspective tends to be voiced by those who have a direct interest in economic matters (governments, business leaders, academics, commentators, and so forth) and at a variety of vantages: from global through to nation states.

For example, business leaders such as Edward de Bono and his book *Serious Creativity* (1992); Tom Peter's *Re-imagine* (2003); or Richard Florida's *The Rise of the Creative Class* (2002), and other best sellers, like Ridderstrale & Nordstrom's *Karaoke Capitalism* (2004), have each emphasized the value of creativity within a global economic context. DeBono sums up the argument:

Cost-cutting, downsizing and quality have run their course. The great need now is for new ideas....what happens when you have a lean and competent organization? What is this lean and competent organization going to do? What happens when your competitors are just as lean and competent as you and your cost-effectiveness is no longer a unique advantage? The more able senior executives know that creativity is now the main hope (de Bono, 1996, Introduction, para. 10).

Politicians too place emphasis on the value of turning bright ideas into businesses, which are "vital for our long-term growth" (Cable & Pickles, 2010, para. 7). Equally, many creativity research papers introduce the value of their research within the context of

economic growth, and in this respect, this part of the thesis follows a well-worn path. Unfortunately, such arguments to longstanding advocates of creativity risk being blasé, possibly rote. Despite 20 years later (a global financial crisis of 2008, and the ongoing exponential growth of digitisation) are de Bono's rhetorical questions to senior executives above any less valid today than back then?

For a new generation of commentators, creativity is a less well-worn buzz word: one positively associated with innovation and discussed alongside science and engineering (Climer, 2016). Yet, creativity can remain exotic or quirky, as expressed through the stereotyped student of Art and Design. Partly this is because the culture of Schools of Art and Design occupies a unique place in academia and university life; neither fitting easily within vocational training nor purely theoretical. It is an education to prepare students for professional creative practice, and at the same time create a space to question the directions and future of that creative practice, and its cultural value. It is for this reason that in the UK, higher education for creativity is esteemed throughout the world, and supports a creative industry that has continued to outgrow the UK economy year on year. Far from quirky, that is a remarkable economic success.

For example, the UK government's 2016 report (DCMS, 2016b) on economic estimates for the Creative Industries highlighted that the Gross Value Added (GVA) by the Creative Industries increased by 6.0 percent each year between 1997 and 2014. A statistic that, compared to the UK a whole, was evidence of the Creative Industries consistently outperforming the UK economy. The report also identified the Creative Industries as accounting for £84bn of GVA -equivalent to around 5 percent of the UK economy. For particular areas of the Creative Industries, growth has been considerably stronger, for example, graphic design had the largest increase in GVA between 2013 and 2014, that of

nearly 17 percent.

Additionally, Nesta (prior to 2010, NESTA: the National Endowment for Science, Technology and the Arts) undertook a comparative analysis of employment within the Creative Industries across member states of the European Union (EU). The key findings (Nathan, Pratt & Rincon-Azner, 2015) highlighted that relative to the EU as a whole, the Creative Industries accounted for 7.58 per cent of employment in the UK compared to 5.21 per cent in the EU, during the period 2011-2013. Regarding the workforce within the Creative Industries, the UK was second only to Germany (3.14 million creative economy workers, compared to 2.94 in the UK). Such findings present higher figures than in previous studies, but the authors account for this through improvements in datasets and more precise classification of creative occupations and industries. The headline-grabbing statement is that the "UK accounts for a fifth of EU's creative industry jobs" (Nesta, 2015). In this context, for economic reasons alone much is required of creativity in relation to academia and industry, and is at stake.

1.4 Outline of Thesis Structure and Chapters

Whether considered in isolation (regarding the development of creativity theory, design education, or economic productivity) or as a synergistic whole, the focus of this thesis on skill, creativity and domain have remained timely throughout its duration. Taking place over six years, the structure of this thesis is defined by five interconnected research studies, presented through nine chapters. Each study builds on the insights and findings of its predecessors, towards the overarching purpose to test: *To what extent are domain competencies correlated with creative output in graphic design?*

Chapter 2 considers literature on how skills, creativity and domain are defined, and in the ways each core term might interact with the others. With definitions and relationships established it considers what methods are appropriate to operationalise these constructs.

Chapter 3 refines the literature review focus with a detailed examination of the Consensual Assessment Technique (CAT). This is presented through the findings of **Study 1**, and the first systematic literature review on the use of the CAT in design research.

Chapter 4: builds upon the challenges of how to prioritise, for research purposes, the skills, knowledge and abilities suggested as important for graphic design. Through **Study 2** it presents a conceptual framework for mapping previous models of graphic design competencies.

Chapter 5: begins to explore the adaption of previous CAT practice, and test the assumptions that the CAT would be a reliable method of assessing graphic design creativity. **Study 3** considers what constitutes a valid and reliable task (or design brief) to be given to participants through two tasks, and finds that a graphic design CAT is more

sensitive to research protocols than had been anticipated. This initial exploration in **Study 3** set a key theme around CAT protocols and became a focus for **Study 4** and **Study 5**.

Chapter 6: extends the findings of **Study 3** through re-examining the challenges the CAT appears to have as a method of graphic design creativity assessment. Specifically, it explores how the CAT needs to be optimised for this purpose, and tests adjustments to CAT protocol through **Study 4**.

Chapter 7: combines all the previous knowledge gained, and in **Study 5**, the final study, addresses the overarching research aim of testing the relationship between graphic design competencies and creativity.

Chapter 8: Is a critical discussion of all preceding chapters, and consider the novelty, merits and limitations of the thesis as a whole, and directions for future research.

Chapter 9: Concludes with a summary of the main findings. A summary of key empirical findings related to graphic design creativity assessment and outline of adjustments to CAT protocol throughout the thesis was as follows:

	Study 3	Study 4	Study 5
Task	Task 1 Task 2	Task 1 Task 3	Task 1 Task 3
Artwork	Prior work	Prior work: Pre-selected	New/Live study
Participants	Novices Intermediates Experts	Intermediates	Novices Intermediates Experts
Judges	5 Novices (Task 1) 5 Experts (Task 1) 8 Experts (Task 2)	16 Experts	8 Experts
Instructions to Judges	Creativity	Creativity Creativity Only	Creativity Only (verbal re-emphasis) Technical Execution Aesthetic Appeal
Rating Sheet	Check List Paper Based ⁴ PowerPoint ⁴	Paper Based	Paper Based
Consensus on Creativity	Task 1: 81%: Creativity ^{1,2} (Novices Judges) 69%: Creativity ^{1,2} (Expert judges) Task 2: 56%: Creativity ³ (Expert Judges)	Task 1: 73%: Creativity ⁵ 82%: Creativity Only ⁶ Task 3: 75%: Creativity ⁵ 80%: Creativity Only ⁶	Task 1: 92%: Creativity Only ^{7,8} 88%: Technical ⁹ 88%: Aesthetic ⁹ Task 3 81%: Creativity Only ^{7,8} 70%: Technical ⁹ 75%: Aesthetic ⁹
1. Is Task 1 a reliable measure of graphic design creativity for both expert and novice judges? 2. For Task 1 how do novice judges' ratings of creativity compare with expert judges' ratings? 3. Is Task 2 a reliable measure of graphic design creativity for expert judges? 4. Can rating sheet design reinforce instructions to use the whole rating scale for Task 2? 5. Can expert judges achieve consensus at or above 70% for Task 1 and Task 3? 6. Do Creativity Only instructions increase consensus on Task 1 and Task 3 7. Without pre-selection of artwork are Task 1 and Task 3 reliable measures of graphic design creativity? 8. To what extent are Task 1 and Task 3 correlated with each other? 9. To what extent are technical execution and aesthetic appeal correlated with graphic design creativity?			

Key empirical findings/paraphrased research questions and the interrelationship to CAT protocols during Study 3, 4 and 5

Chapter 2: Definitions and Relationships Between Domain, Skills and Creativity

2.1 Introduction to Chapter 2

In this chapter, each of the key terms (domain, skill and creativity) will be explored. The first focus will be on defining these terms, and what methods are available to identify them: for example if creativity is defined as a creative output, then what methods are established on which to assess creative output?

The second focus will be on the relationships between these key terms; for this purpose, a Venn-like diagram is proposed as a way to manage the complexity of these relationships and offer a means to map the research literature. The diagram (Figure 1) highlights that the key relationships to be explored will be as follows, between creativity and domain; domain and skill; skill & creativity.

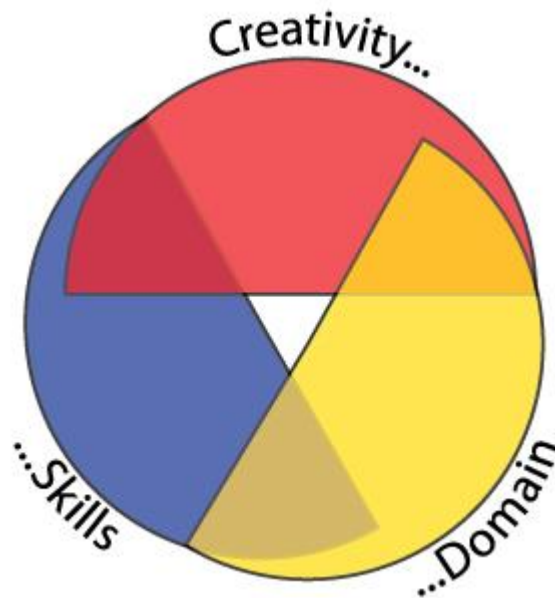


Figure 1: Key relationships between skill, creativity and domain

At the core of this diagram is where research on skills, creativity and domain overlap. Through initially separating these relationships, it is hoped that the nuances and later analysis of works that fuse these three terms together will be clearer to understand for the reader. The purpose is to place the thesis within a broad research context and to give the background required for certain research designs and methods to be discussed throughout the thesis. A useful point of departure is first to define, and identify, what is meant by domain.

2.2 Defining Domain

Creativity occurs when a person, using the symbols of a given domain such as music, engineering, business, or mathematics, has a new idea or sees a new pattern, and when this novelty is selected by the appropriate field for inclusion into the relevant domain. (Csikszentmihalyi, 1996, p.28).

The quote above highlights that for some researchers, the notion of a domain is directly related to identifying creativity. Equally, other factors are inferred: knowledge, skill, expertise. For Csikszentmihalyi (1999) an individual's creative contribution is only one element within a system made up of the domain, the field, and the individual. Within each domain is a body of tacit and explicit knowledge, and it is this knowledge that becomes the reference point for the criteria of newness and originality. The field refers to people responsible for a domain: educators, practitioners, professional organisations, critics, researchers, historians, and so forth. Lastly, the individual: the actual person (or persons) who creates the output. Csikszentmihalyi (1996) considers the intersection of domain, field, and individual to profoundly influence identifying creativity. Indeed, in this systems view of creativity, it is the field that confers the title "creative." As shall be discussed, such theories are highly controversial (Runco, 1999; Nickerson, 1999). However, whether or

not researchers subscribe to social recognition being fundamental to identifying creative output, creative output often does occur within or aims to break away from established domains. Thus, it is important to understand more specifically what is meant by domain.

2.2.1 Definitions of Domain

In this regard, research offers a broad spectrum of what is meant by domain. This can extend from what has been described as domains of mind (Feist, 2004) through to specific occupational roles within established industries (Skillset, 2003), and for some researchers, this culminates at the level of microdomains: poetry writing, short story writing, etc. (Baer, 1998).

2.2.2 Domains of Mind

Theories based on domains of mind (Feist, 2004) propose that distinct conceptual domains exist which have evolved from our survival needs. Broadly, these domains map to core areas of the sciences (psychology, physics, biology), arts (linguistics, music, aesthetics), and mathematics (numerical). The inclusion of such domains is argued to be justified by, among other evidence, research from archaeological, genetic, and psychological studies.

2.2.3 Occupational Roles

Other researchers consider the concept of a domain as shorthand for describing any craft, discipline, art, sport, or occupation (Gardner, 2006). This broad definition of domain is interesting as this closely aligns itself with a significant body of work related to National Occupation Standards and the work, in the UK, of the Sector Skills Councils. Creative Skillset, for example, in addition to developing National Occupational Standards, also offers job profiles for the audio-visual industries (Skillset, 2005; Skillset, 2004b), and these

give a framework within which to consider an occupational taxonomy for specific domains. Indeed, such documents clearly highlight the breadth and complexity of the audio-visual industry through an extensive A to Z of job roles from first assistant director to wardrobe master/mistress.

2.2.4 Microdomains

More detailed and highly specific definitions of domain are present in the concept of microdomains (Baer, 1998) and work where domain and specific task undertaking begin to merge (Baer, 1993). As mentioned earlier, this research considers creativity at the level, for example, of short story writing rather than the wider domain of creative writing.

2.2.5 Which Definition to Use?

The implication arising from this breadth of definitions of domain is which definitions of domain to use: domains of mind (the visual arts); broad occupational domains (audio-visual industry); related occupational domains (types of film directors); single occupation roles (director for corporate film); or, at the level of specific tasks (for example, when a film director develops initial ideas for a film). Baer (1998) would argue that in the absence of a conclusive answer to this question, it is wise to side with defining domain at a fairly specific level. The rationale for this is interwoven with the debate regarding whether creativity is domain-specific or domain-general. However, if such an argument implies that domain is defined at the level of specific tasks, as aspects of Baer's (1993) research have previously, then this presents significant operational challenges for creativity research: these are studies at an extremely detailed level.

2.2.6 Related Occupations

Jeffries (2007b) argued for a compromise between defining domain at too broad or too

specific a level and advocated defining domain at the level of related occupations (for example, directors for film, video, television), but proceeding with caution. It would seem prudent that those researching related occupational domains be aware that their findings may be biased towards generalisation. However, the benefits of such work could give further insight into the domain-generality/specificity of creativity. Indeed, this type of approach may be able to identify consistent themes across occupational domains or conversely, show that even within related occupational domains little consensus exists, implying the need for single occupation or task specific studies.

2.2.7 Single Occupations

A key finding from this study of the audio-visual industry (film, television, etc.) was that the inclusion of related occupations was problematic for some domains, and not for others (Jeffries, 2007b). The explanation given was that the occupational domains which formed the visualiser group were too diverse, particularly in comparison to the director and sound group. While participants in the visualiser group could identify themselves as an artist, art director, standby art director, or graphic designer, their involvement in specialist areas within these single occupational was possible. For example, art directors may specialise in one area or work across several: static, moving, two-dimensional, or three-dimensional image.

2.2.8 Domain Identity of Participants: Designer or Artist

As a result of this complexity, subsequent studies have defined domain as the level of a single occupation (such as a games designer). Even at this level, however, taxonomies emerge between participants who work as games designers but would consider themselves games artists (Jeffries, 2011; Jeffries, 2009). Indeed, as previous research has highlighted (Gluck, Ernst, & Unger, 2002), such domain identities may be substantial,

and be suggestive of the distinction between free artists (e.g. painters) and artists in more constrained professions (e.g. designers); who have different perspectives on creativity. With games design, however, it appears that at this level of taxonomy (games designer relative to games artist) a mixed group can achieve consensus. (Though a caveat to consider is that while the group was mixed the focus of the research remained about a single occupation and domain: games designer and games design).

2.2.9 Definition of Domain in this Study

As such, sampling considerations are likely to be present in a study on graphic design.

Some graphic designers may consider themselves graphic artists; others may prefer the term visual communicator; it is also possible that typographer could be a term some experts in graphic designers would prefer. Thus, the position taken in this study is to define domain at the level of a single occupation (graphic designer) and to control the sampling to include experts who identify themselves with this domain (but who may describe themselves as a graphic artist, visual communicator).

A further consideration in relation to domain is how much time a person spends in their occupation, for example, do they work part-time, or full-time; indeed, the self-employed nature of particular professions can mean full-time work is intermittent. As the previous study on the audio-visual industry highlighted, some freelancers use the time between film projects for employment within other areas of the audio-visual industry, and some use the time to pursue their personal artistic interests. How participants use this time between projects may influence their values about creative output. Whilst, it could be argued that intermittent full-time work is a particular feature of the film industry, the point of divided interests could equally apply to part-time work in many other professions. It is anecdotal, but not uncommon to find individuals with loyalties to a particular domain (be it acting or

music, as examples) working to "pay the bills" through work within another domain. To account for this, experts will be sampled from graphic designers working on a full-time basis.

A final point to consider is why choose graphic design, why not games design or film direction as the previous studies above?

Within the Creative Industries graphic design links across many sectors: be it the need for marketing material; a new logo for an organisation; the presentation of scientific information, or the development of a new product. Graphic design will play a part: sometimes in the background, at other times centre stage.

Established within traditional print media alongside the development of the Digital Creative Industries, the roots of graphic design, arguably, do not first begin in the 20th century. Visual communication: turning words into icons, or using images *to say a thousand words* is ancient. Like many established professions its boundaries blur and change, and the line between graphic design, illustration, advertising or copywriting is debatable and in flux (we could add other domains to this list: typography, printing, web design, photography, animation, packaging, let alone new technologies searching to establish an identity).

Yet, within this flux, graphic design exists and is global. Throughout the world, universities, colleges and private companies run graphic design courses; their students' gain qualifications that enable them to work for design agencies or set up their own business, and they in turn work for clients who require graphic design to appeal to their audience and customers. In this context, creativity is an asset and gives clients, design agencies, and individual graphic designers an edge in a competitive market. Whilst not all clients

and educational opportunities give scope for graphic design creativity, many do.

Moreover, the type of creativity required for graphic design can, depending on the task, still be the achievement of a single designer. In terms of empirically testing the relationship between skills and creativity, this has a number of advantages for research design. This is not to discount the role graphic designers play within multidisciplinary design team, nor to infer that graphic designers need to work as a silo, nor to underplay the value of research related to team creativity; it is to state that compared to the complexity of domains involved in the production of a digital game or film, graphic design (at the outset of this research study) appears to offer the opportunity to control the social interactions involved in the creation of a final design without undermining the "real world" relevance of the creativity being studied.

2.3 Defining Skills

Within various literature, it can be fraught to isolate skills from domain when considering creativity. For example, building upon Amabile's componential model of creativity (Amabile, 1983), domain relevant skills have been described as the skills, knowledge, or talents required for competent performance within a domain (Conti, Coon, & Amabile, 1996). Amabile, in other works, gives the example of when a composer creating a new musical composition is able to hear the sound of the instruments playing together in his or her imagination (Amabile, 1996); such a "talent" may be crucial to music composition or trivial, "...only future research can indicate...which elements are indeed essential" (p. 85).

2.3.1 Conflated Usage

What is immediate, in both these examples, is that the term "skills" is conflated with "knowledge" and "talents": such conflation is not an isolated occurrence. Indeed, as

competency is also interwoven within such discussions, it is fairly common to find knowledge, skills, abilities, beliefs, values and interests being presented as synonyms within other works in this area (Fleishman, Costanza, & Marshall-Mies, 1999).

Moreover, organisations, such as the Council for Industry and Higher Education, write about the need for "Hybrid skills", of which creativity, along with technical, business, and interpersonal are presented as clarifications for this term (CIHE, 2010). A similar theme can be found in other works. The Design Skills Advisory Panel (Design Skills Advisory Panel, 2007) discussed the crucial value of "High-level skills" in design, stating that:

If the UK is to increase its capacity for innovation, then it needs an education system that is able to develop foundation-level analytical and problem-solving skills, plus creativity, imagination, resourcefulness and flexibility. These are the skills that design education can provide for everyone. (p.31).

Levy and Hopkins (2010), continue with the focus on high-level skills (occasionally described as "Higher Level Skills") and define that: "...the notion of high level skills is a broad concept reflecting an individual's ability to use tacit knowledge to assimilate and interpret information" (p.32).

Clearly, nomenclature and taxonomy of terminology changes depending on which literature is cited: some describe skills along the lines of abilities; others with knowledge (either explicit or tacit knowledge), talents, or creativity itself. It is beyond the scope of this review to do justice to the breadth, complexities, and contradictions regarding the usage of the term skills. The purpose of this section is to define skills in a manner that is relevant to this study and, in doing so, explain the rationale for why some areas of research

literature have been favoured over others.

2.3.2 Talent and Giftedness

Talent, for example, has been defined as "a constellation of inherited skills that make a person especially suited to excel in a specific domain" (Weisberg, 2006, section Scepticisms about Expertise and Creativity, para. 1). Of particular interest here is the notion of inherited skills, as a controversy with research on talent surrounds whether or not talent is innate. A similar issue needs to be considered for research within the area of giftedness. In addition, researchers have also highlighted that the field of giftedness still needs to form its own definitions of what it means by giftedness, and as such continues to be disparate and complex at the present time (Dai, Swanson, & Cheng, 2011). For these reasons, research on talent and giftedness has been limited in this review, though clearly, they have much to offer our understanding of creativity and domain.

2.3.3 Expertise

Research on expertise is a well-established field (Ericsson, Charness, Feltovich, & Hoffman, 2006), and has direct links with knowledge, skill and competency, in addition to creativity. Indeed, a number of seminal academics within creativity research also produce works on expertise. Such research has developed conceptual frameworks and definitions, and of particular interest to this study is work in the area of vocational psychology, and within this, job analysis.

In this respect, the work of Fleishman and colleagues (Fleishman, Costanza & Marshall-Mies, 1999; Fleishman & Reilly, 1992) is of particular note, and specifically their development of the F-JAS (Fleishman Job Analysis Survey). It was the F-JAS that became the basis for aspects of the Occupational Information Network (O*NET) which

was developed to gather data on all jobs within the United States and to identify the types of competencies required for specific job roles (O*NET Online, 2016).

O*NET has been described as a comprehensive system that "incorporates the last 60 years of knowledge about the nature of jobs and work" (Peterson et al., 2001, p.451), and a key aspect of its development was to clarify common language on terms and a hierarchical taxonomy with which to navigate from the broad to the specific. As a result, it draws a number of conceptual distinctions between, skill, knowledge, ability, tasks and competency.

2.3.4 Competency

As mentioned above, competency can be vague in usage, however, as a broad concept, it describes a level of performance in relation to a particular job: such performance ranging from effective to superior (Boyatzis, 2008). In this study, the term competency is used in its broadest usage; to denote an overarching concept under which skills, abilities and knowledge are related to effective performance in graphic design.

2.3.5 Abilities

Related to competency is ability. As a term, its precise conceptual definition has been debated in relation to whether it is a trait, something that can be applicable across a number of domains, or amenable to development (Fleishman, Costanza & Marshall-Mies, 1999). They would argue ability is a general trait, and as such remains present over time, whilst acknowledging that it is feasible to develop ability -though again this occurs over time. Furthermore, abilities tend to be underlying traits, and tend to form the basis on which skills can be developed. Thus, for this study, abilities are seen to mediate between

competency (at the broadest level) and skills (at more specific levels). However, abilities, like skills, can be categorised too broadly and therefore have little value for research purposes, or too narrowly and equally obscure the underlying abilities that are required. Via Fleishman's research on types of abilities, the O*NET content model identifies four core abilities: Cognitive, Psychomotor, Physical and Sensory, which cascade down to 52 lower-level abilities, such as written expression, finger dexterity, stamina and depth perception, for example.

2.3.6 Skills

Skills, in this respect, appear to be conceptually distinct from abilities (Peterson et al., 2001). Skills can be acquired quicker than abilities, and they are more amenable to improvement via direct training and education. How effective such skills training or education is, can be dependent on the type of abilities that may underpin a skill. As a general example, driving skills can be improved by training if requisite underpinning abilities are present: these may include abilities with spatial awareness, or hand to eye coordination, for example. O*NET's content model distinguishes between basic skills (that facilitate learning) cross-functional skills (that facilitate problem-solving across occupations: for example, negotiation skills) and occupation-specific skills. In this respect, O*NET is similar to other areas of psychology which distinguish between domain-dependent skills (for example, chess playing) and domain-independent skills (for example, solving simple puzzle problems) (Green & Gilhooly, 1992; Gilhooly & Green, 1989).

2.3.7 Tasks

By highlighting the issue of domain-dependent skills and domain-independent skills, the terminology raises the issue of domain-specificity and domain-generality (as mentioned in the Introduction: theoretical controversies within creativity research, and pedagogic

concerns between academia and industry). Within O*NET's content model, tasks form a fundamental basis on which to consider the types of specific skills, abilities and knowledge required for a job, and by extension competency. It identifies occupational tasks via subject experts' consensus on general work activities and essential tasks for their occupation. In this respect, O*NET's version 16.0 job analysis for a graphic designer highlights the top five core tasks of a graphic designer and does so in order of importance. These findings are discussed in detail in section 2.6.1 (O*NET Graphic Design) of this thesis.

2.3.8 Knowledge

Knowledge, in this model, is defined by (Costanza, Fleishman, & Marshall-Mies, 1999) as a "collection of discreet but related facts and information about a particular domain" with the caveat that how such information is "organised into some coherent structure is critical to the definition" (p71). Defining knowledge in this way appears to emphasise what could be termed explicit knowledge: facts and information that can be codified, written down, or put into words in some manner. In contrast, the discussion of tacit knowledge may appear absent from this perspective (Sternberg, 2006; Sternberg & Lubart, 1995). Indeed, from the work of Michael Polanyi (Polanyi, 1958) through to Harry Collins (Collins, 1985; Collins, 2010), discussion surrounding tacit knowledge is a complex area of philosophical debate; some philosophers argue that tacit knowledge is practical knowledge (know-how) but remains limited in how it can be transmitted through words (Thornton, 2012). For this study, knowledge can be explicit, tacit, and practical, it can also be specific to a domain, or apply more generally across domains. Moreover, the coherence of how this knowledge is structured is a consideration and may relate to links between creative output and whether knowledge is retained as conceptual abstractions and principles or coded as prescriptive facts and rules (Ward, Patterson, & Sifonis, 2004). As above, like tasks,

abilities and skills, O*NET's version 16.0 job analysis for a graphic designer highlights the knowledge required, and will be discussed in detail in the section on domain and skill.

2.3.9 Definition of Skills in this Study

Part of the value of O*NET's taxonomy is the clarifications it offers regarding the conflated usage of the term skills: skills are distinct from abilities, and from knowledge. In this context the definition of high-level skills (discussed above) as "...a broad concept reflecting an individual's ability to use tacit knowledge to assimilate and interpret information" (p.32) appears understandably confused.

Arguably, what Levy and Hopkins are describing is more akin to ability, and suggestive of cognitive ability, possibly one to be classified under Idea Generation & Reasoning. In this respect, O*NET's definition of Inductive Reasoning seems apt: The ability to combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events). If for example, higher level skills is another name for inductive reasoning ability then, from a research perspective at the very least, this is useful to know. It makes operationalising such a construct more feasible; builds upon existing research, and enables comparison of findings across studies. Such distinctions are important for at least three other reasons.

Firstly, abilities underpin skills; without requisite abilities the opportunity to develop skills for particular tasks is limited. In this respect, abilities are fundamental to skill acquisition and task competence. To state this hierarchy of importance is, in itself, useful. Indeed, the lack of such hierarchy may partly explain why debates on design skills, and skills more generally, in the UK have suffered from such conflated usage and challenges to test construct validity. Notwithstanding, whilst empirical studies on design abilities are likely to

remain complex, theoretical models do exist which offer a comprehensive framework to focus current research studies, and build upon past research findings. Fleishman, in his preface to the Handbook of Human Abilities (Fleishman & Reilly, 1992), makes a similar point on the future direction for F-JAS, that researchers need to use frameworks that build upon the existing research on human abilities: of which a substantial body of knowledge has already been developed (Peterson et al., 2001). Moreover, when significant organisations and researchers develop nomenclature that evolves in isolation to established theoretical frameworks the likelihood is that research findings get lost in their own complexity, or do not identify constructs well researched within related fields (in this case, abilities research and job analysis vocational psychology). In essence, at the current time more is to be gained from a shared language and theoretical taxonomy than is lost.

Secondly, the design industry needs designers who have both the abilities and skills that underpin tasks competency. But, of these two, skills can be acquired (with the requisite abilities) more readily, and are amenable to short-term training. The reality of many design professions is that the skills required today can quickly be redundant in the light of new technical developments: upskilling is thus inherently unstable and requires a continuous process of professional development for designers and the design industry. The extension of this point is to state that the onus, then, for skills development is best placed with individual designers and their employers.

However, thirdly, such a statement has implications for higher education. Abilities are acquired over longer timescales than skills, and as discussed, need to be in place in order for skills acquisition to be effective. Abilities, clearly, continue to evolve throughout a designer's professional development, but in order to initially engage, and then keep pace with their profession, certain abilities need first to be in place. For the most part the onus,

then, is on design education and, more specifically, design higher education to provide graduates with a climate in which the relevant abilities can flourish. The caveat is that most UK undergraduate degree courses in design take around three years of full-time study, and most postgraduate degree courses take one year of full-time study. Depending on the abilities required, four years may still not be enough time. If this scenario occurs then abilities developed during secondary and primary education will form a fundamental basis upon which design higher education will need to build.

Though the question remains what actual abilities are required to undertake graphic design at a professional level, the conclusion of this section is to state that the definition of skills used in this study will be based on O*NET's conceptual framework. And, whilst the use of the term skills in various literature is conflated with other constructs, for this study the focus is on abilities.

2.4 Define Creativity

In order to define creativity, and operationalize creative output, it may be useful to first address the broader concept of creativity in relation to creative output.

2.4.1 Eastern and Western Definition of Creativity

As the eminent creativity researcher E. Paul Torrance (Torrance, 1995) highlighted, a familiar and cyclic question is "What is creativity?" Research on this topic has generated a wealth of literature and depth of debate (Runco, 2004), and one of the first distinctions to highlight is that much of this literature has a particularly Western perspective on creativity. Over the last two decades, creativity research has become more aware of how creativity is conceptualised in Western cultures, often in contrast to Eastern Cultures (Batey & Furnham, 2006). The implication has been that certain creativity research methods

accepted in previous eras are now seen to be more limited in their global application than was first assumed. As the literature in this thesis is, for the most part, premised on a Western outlook, either explicitly or implicitly within the studies cited, then these findings may not apply outside of this cultural context.

2.4.2 The 4 P's of Creativity Research: Product, Process, Person and Press

Premised on this context, creativity researchers consider creative output as one of four means to define creativity; other categories of research explore the creative process, the creative person, or the creative press (environment): known as the 4 Ps of creativity research (Isaksen et al., 1993). While creative output has occupied a considerable focus for researchers, some researchers do have an unease with creative output as a measure of creativity. Indeed, many creativity researchers involved in kindergarten and primary education consider creative output a rather brutal operational definition of creativity in those settings.

2.4.3 Historical and Personal Creativity

With this in mind, the consensus, though, is that creative output shows both the qualities of being original and useful (Mayer, 1999). Such a definition emphasises the tangible outputs of creativity: a painting, a music composition, a scientific discovery, etc. Moreover, scholars draw a distinction between historical and personal originality (Boden, 1994), proposing that personal originality is specific to the viewpoint of the creator, for whom the product created is utterly new. From a historical perspective, a product may be new to an individual creator but not new within the context of human history. This distinction can be controversial both in how we define core characteristics like originality and the type of evidence required to assess creative outputs (Runco, 1999).

In relation to design, it has been suggested that design education tends towards a pragmatic view regarding assessing creative outputs (Jeffries, 2007a). Ideally, the outcome of a design education is that students produce historically original work; ultimately, designs that are fresh and new to the domain (and useful with respect to answering the brief). Pragmatically, though, in terms of educational development, there is a hierarchy of attainment: achieve personal originality first then build toward historical originality with educational progression.

2.4.4 Creative Output

An argument can be made (for some creativity researchers this argument is generally accepted) that to validate findings in these three other areas of creative process, person or press, verification needs to be shown through creative output at some point (Kaufman & Baer, 2002). By doing so, this places the primacy of creative output at the heart of creativity research and creativity assessment. This does not necessarily mean that those that have no creative outputs are uncreative, more so that for research purposes, creative output is a useful definition. From this perspective, the measurement of creative output, with adult participants familiar with the assessment of their creativity, seems less ethically problematic and contentious than with children or adolescents.

From such a position, then, the focus can move towards broad methods with which to assess creative output. As these are considerable and varied (Batey & Furnham, 2006), four areas will be discussed: protocol studies, self-report measures, divergent thinking tests, and creativity assessment by domain experts, with specific reference to the Torrance Tests of Creative Thinking, and Amabile's Consensual Assessment Technique.

2.4.5 Protocol Studies

The absence of protocol studies from this review should be noted; particularly in relation to creativity studies in design, this has been a favoured method. Protocol analysis (Ericsson & Simon, 1984, 1993) enables researchers to explore the type of cognitive processes and decisions participants make during an activity, for example when sketching ideas for a design. Such study is crucial to our understanding of the creative process, but, as argued above, this study will be focused on the creative output rather than the creative process. The purpose is not to understand the creative processes during a task (important as this is), but to test the relationship between creative output and domain skills within graphic design. For this reason, Protocol studies have not been part of this review.

2.4.6 Self-Report Measures

For different reasons, the use of self-report measures has been limited in this review. Again, this is a valuable and established approach to creativity research (Lau, 2011; Batey, Furnham, & Safiullina, 2010). Some researchers have argued that given the emphasis within design education for assessing creative outputs, data gathered by self-reports offer an alternative perspective on creative potential (Jeffries, 2007a; Kaufman & Baer, 2002). Self-report data enables researchers to examine an individual's self-image in relation to their creativity, and such factors are important for theoretical and pedagogic reasons. There are, however, acknowledged challenges to the use of self-report instruments such as, not being easily verifiable, and open to 'halo' effect bias on the part of the participant (Brown, 1989; Lubart & Guignard, 2004). It is for these reasons that self-report measures have not been part of this review.

2.4.7 Divergent Thinking Tests and the Torrance Tests of Creative Thinking

Few would argue that E. Paul. Torrance's impact on creativity research was anything but remarkable. It is no small measure of this researcher's esteem within the field that the

Creativity Research Journal dedicated a whole edition to his work and legacy after his death. Of his many accomplishments, he developed the Torrance Tests of Creative Thinking (TTCT), (Torrance, 1966; 1998). As the first prototype of its kind, in subsequent years the TTCT became the forerunner for many other tests developed to measure creativity. It has been used in a vast number of studies over the years, and as a result, it has its advocates and critics. TTCT is a type of divergent thinking test that presents a number of open-ended challenges that require a visual (sketch) or a verbal (written) response, and these responses are assessed on a number of measures based upon, for example, the number of responses given (fluency), or the originality of the response relative to other standard responses: such standards change depending on age.

2.4.8 Creativity Assessment by Domain Experts, and the Consensual Assessment Technique

Teresa M Amabile's Consensual Assessment Technique (CAT) has been described as the "Gold Standard" in creativity assessment (Kaufman, Plucker, & Baer, 2008). It has been used extensively in creativity research since first introduced in 1982. By emphasising the subjective element in creativity assessment, Amabile's CAT is at the opposite end of the assessment spectrum to TTCT, and to appreciate this it is helpful to give some background history regarding the development of the CAT, in relation to TTCT.

Although a significant body of work had been completed on creativity, the majority of studies, prior to the 1980's, had focused on the psychology of individual creativity. Amabile (1982) argued that you couldn't truly understand creativity without taking the social context of creativity into account: relationships with others, particular environments, externally imposed working constraints, etc. When Amabile began her enquiry into the social dimension of creativity she was faced with a research design problem: many of the accepted research methods were not appropriate for social studies on creativity, for

example, the TTCT was not applicable to this type of study focus.

There were a number of reasons for this. The TTCT and other divergent thinking tests are about quantifying those aspects that set people apart; they aim to define and quantify the micro and macro factors that distinguish one individual from another. In contrast, researching the social context of creativity requires the need to define and quantify group characteristics beneficial for comparing a control group against a test group. Thus, there was a need to minimise individual difference, in order to test hypotheses about how one group may react differently to another, given changes in their social environment. The lack of an acceptable research methodology meant a new method to assess creativity needed to be evolved, and to prove its validity. This was the purpose of the CAT.

If we accept the basic definition of creativity mentioned, that creativity produces work that is both new and useful, we could state that our initial criteria for assessing creativity could centre on assessing how new and useful is the final product. From such a position we are left with several questions: what are the appropriate assessment criteria for new and useful? How should assessors evaluate new and useful, and how do they do this with transparency and objectivity. Amabile argued (1982) that objective methods did not currently exist (and may never exist) on which to assess creativity in this way. Moreover, that the judgements required to assess creativity "...can ultimately only be subjective" (p. 1001). And as such, with an appropriate group of judges, "...is something that people can recognize when they see it" (p.1001). From this came an operational definition of creativity, upon which the CAT is based. That is, "...a product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated" (p.1001).

Using this as a working definition Amabile moves away from the notion of objectivity in assessment and towards subjectivity. The question then is to what degree can judges actually agree on each other's subjective opinions?

By basing the criteria on the judges' subjective opinion, this also negates the need for explicit criteria. As long as the judges are in agreement, then that is enough; they may not know specifically why a product has a certain level of creativity, but if they agree that it does, then this shall form the basis for evaluation. It is worth highlighting that, for these reasons, the CAT is considered to be a theoretically neutral measure of creativity, in that suitable judges are free to use whatever tacit or explicit criteria they may have evolved regarding creativity within their domain. In this respect, the CAT is quite distinct not only to the TTCT but other measures of creativity that define the criteria judges should apply when rating creativity: for example novelty, surprise, or usefulness. This theoretical neutrality is one reason creativity researchers have advocated the CAT as a “Gold Standard” of creativity assessment (Kaufman, Plucker & Baer, 2008).

Given such a radical departure from assessment norms at the time, the Consensual Assessment Technique was developed to evaluate whether such levels of agreement actually exist and to what degree they are reliable and consistent. Over a five-year period, Amabile conducted eight studies using the CAT. With a wide range of groups represented from primary, secondary, and undergraduate education: the total number of students engaged in the research was 423. These groups either took part in a study to assess visual or verbal creativity. A range of assessors recruited from academia, working practice and education judged this work. The total number of judges over the five years was 125; each judge was free to use his or her own subjective definition of creativity with which to assess the work. From these studies, Amabile concluded that high levels of judge

agreement existed on the creativity rating, with the results showing significant reliability when using the CAT. Furthermore, judges were able to distinguish 'creativity' from other aspects of the work such as aesthetic appeal and technical execution.

With these findings in place, the CAT was used as the basis for research into the social impact on creativity. It is beyond the scope of this review to detail the findings from this research, but a number of studies found negative relationships between external evaluation and surveillance on creativity (Amabile, 1979; Amabile, Hennessey, & Grossman, 1986; Hennessey, Amabile, & Martinage, 1989; Amabile, Goldfarb, & Brackfield, 1990). While such findings are disputed (we are now aware of several caveats regarding these early works), the value of the CAT as a research method has continued to evolve. In more recent years, creativity researchers (Baer, Kaufman, & Gentile, 2004) have extended the CAT to less stringent experimental conditions than Amabile and others initially used. Such studies suggest satisfactory results could be achieved with less than 13 judges. Kaufman, Plucker and Baer's later works (2008) have stated that "for most purposes 5-10 judges is an adequate number" (p.74). Equally, Kaufman, Baer, Cole & Sexton, (2008) have explored the use of non-expert raters for the CAT, and find that the requirement for expert judges still holds.

2.4.9 Definition of Creativity used in this Study.

As a result of this type of work, in this study, the CAT has been taken as the conceptual definition of creative output. Where TTCT is premised on the importance of divergent thinking skills (both in the tasks set, but more importantly in the scoring of responses) the CAT is not conceptually tied to any particular skill, knowledge or ability. Theoretically, the only CAT criterion is what domain expert judges consider important to creativity in their domain (be this based on explicit, tacit or practical knowledge). Furthermore, as other

studies have shown the CAT is highly adaptable to assessing creative output within many domains. In theory, it should be adaptable to assessing creativity in graphic design.

2.5 Key Relationship: Creative and Domain

As mentioned in the introduction to this thesis, the relationship between creativity and domain has for years polarized the world of creativity research, and still does so (Plucker, 1998; Han & Marvin, 2002; Plucker & Beghetto, 2004; Silvia, Kaufman, & Pretz, 2009; Reiter-Palmon, Illies, Cross, Buboltz, & Nimps, 2009, Baer, 2012). The reasons why are explored below.

2.5.1 Domain-General Theory

Domain-general creativity theorists suggest that the major attributes required to be creative in one domain can be equally useful within other domains. For example, divergent thinking (the ability to generate a large number of ideas) has frequently been equated with creativity (Cropley, 2006), and as domain transcending, a prerequisite for creativity regardless of domain (Diakidoy & Spanoudis, 2002). Whether in the sciences or the arts, divergent thinking tests form a significant predictor of creative potential (Plucker & Renzulli, 1999; Plucker, 1999). Arguably, many creativity training programmes have been underpinned by domain-generality.

2.5.2 Domain-Specific Theory

In contrast, advocates of domain-specificity propose that creativity is fundamentally influenced by the domain in which it occurs. Creative capability in one domain does not assure creativity in another (Feist, 2004). While motivation, for example, may appear a general requirement for creativity, an individual's motivation to write a short story need not follow the same cognitive process required to write a mathematic equation; they are likely

to be different directly because of the variation in domain (Kaufman & Baer, 2002), and this difference may matter when it comes to enhancing and teaching creativity (Baer, 1996). At the extremes of domain-specificity (Baer, 1993), certain researchers advocate creativity training to be highly specific (tailored to writing a poem, rather than creative writing per se).

2.5.3 Hybrid Models

Whilst, several theories have attempted to bridge this polarity by hypothesising that both the domain-general and domain-specific are important, such models are also seen as compromises (Silvia et al., 2009) that, with the best intentions, sidestep the core debate: to what extent is creativity domain-specific? Nor is the interaction between the domain-general and domain-specific fully understood to suggest which form of hybrid model to pursue: mostly domain-specific; 50/50 domain-specific/domain-general; mostly domain-general.

2.5.4 Conclusion

For the purpose of this thesis, the choice was taken to build upon the assumption that creativity may be domain-specific. The rationale was two-fold.

Firstly, consideration was given to future proofing the results of this thesis, in the event that a conclusive answer to the domain-specific/generalizability of creativity debate is accepted. If creativity appears to be domain-general, i.e. that the research and knowledge developed in one domain are transferable to another, then domain-specific findings, in theory, should still be useful and offer examples of underlying principles. In contrast, if creativity is domain-specific or a hybrid, then domain-general findings could be limited. On these grounds, a cautious stance can be justified toward domain-specificity.

Secondly, having selected the CAT as the operational measure of creative output, this conceptually aligns this thesis towards domain-specificity. While the CAT does not stipulate precise criteria judges must use to assess creative outputs (and in this respect is accepted as the most theoretically neutral methods of creativity assessment), the need for domain expert judges does tend to suggest the CAT more as a domain-specific measure of creativity.

With such a base established, that is, graphic design as the level of domain-specificity to be explored, and CAT as a suitable measure of domain creativity, one key question to understand is what is the relationship between the CAT and graphic design creativity research? For example, how has the CAT been used in this domain, and what guidance can be found in such studies for this thesis. As will become clear shortly (and a broad discussion raised in Chapter 8) the answer at the time of this study was not known. This resulted in the first systematic review of the CAT in design (Jeffries, 2012); the details of this research are presented in Chapter 3.

2.6 Key Relationship: Domain and Skill

As the purpose of this study is not to undertake new research to identify domain competencies, but rather to test existing competencies and their relationship to creative output (if any), then a credible model is required that claims to have already identified what these competencies may be.

The challenge is that many models claim credibility, each with differing levels of detail and theoretical underpinnings. The section below highlights a selection of well-known and established sources for competencies in graphic design, namely: O*NET (Occupational

Information Network; NOS (National Occupational Standards); specific BTEC (Business and Technology Education Council) qualifications relevant to graphic design, and NASAD (The National Association for Schools of Art and Design).

2.6.1 O*NET Graphic Design

The work of O*NET in the United States (US) has been described as "the nation's primary source of occupational information...(and) a set of valuable assessment instruments for workers and students looking to find or change careers." (O*NET Online, 2016). In its development and subsequent revisions, it has engaged a number of seminal academics within the field of vocational psychology. As a result, from the outset, its aspirations were ambitious: to create a major framework, with the aim to identify the prerequisite competencies for every job available in the US labour market. For the purposes here, the focus is on O*NET research on graphic design.

O*NET's work suggests specific tasks, skills, knowledge and abilities for graphic design, and hierarchy, i.e. the most important skills for a graphic designer, presented in order of importance. It is this combination of detailed analysis with prioritisation that makes this work of particular value from the research perspective of this study. At the outset, it appears to offer validated constructs that will be amenable to empirical study. In this respect, O*NET's version 16.0 job analysis for a graphic designer highlighted the top five core tasks of a graphic designer, and the following skills, knowledge and abilities are suggested in order of importance (0-100), shown in Table 1, 2, 3, and 4.

*Table 1: O*NET's occupation-specific tasks for graphic design*

<ol style="list-style-type: none"> 1. Create designs, concepts, and sample layouts based on knowledge of layout principles and aesthetic design concepts (98); 2. determine size and arrangement of illustrative material and copy, and select style and size of type (94);

3. confer with clients to discuss and determine layout design (85);
4. develop graphics and layouts for product illustrations, company logos, and Internet websites (83);
5. review final layouts and suggest improvements as needed (81).

*Table 2: O*NET's abilities for graphic design*

1. Originality - The ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem (72)
2. Written Comprehension - The ability to read and understand information and ideas presented in writing (66)
3. Fluency of Ideas - The ability to come up with a number of ideas about a topic (the number of ideas is important, not their quality, correctness, or creativity). (63)*
4. Near Vision - The ability to see details at close range (within a few feet of the observer) (63)*
5. Written Expression - The ability to communicate information and ideas in writing so others will understand. (63)*

*(Note 3, 4 and 5 have tied scores)

*Table 3: O*NET's knowledge for graphic design*

1. Design - Knowledge of design techniques, tools, and principles involved in production of precision technical plans, blueprints, drawings, and models. (92)
2. Communications and Media - Knowledge of media production, communication, and dissemination techniques and methods. This includes alternative ways to inform and entertain via written, oral, and visual media. (78)
3. Computers and Electronics - Knowledge of circuit boards, processors, chips, electronic equipment, and computer hardware and software, including applications and programming. (72)*
4. English Language - Knowledge of the structure and content of the English language including the meaning and spelling of words, rules of composition, and grammar. (72)*
5. Fine Arts - Knowledge of the theory and techniques required to compose, produce, and perform works of music, dance, visual arts, drama, and sculpture. (68)

*(Note that 3 and 4 have tied scores)

*Table 4: O*NET's skills for graphic design*

1. Active Listening - Giving full attention to what other people are saying, taking time to understand the points being made, asking questions as appropriate, and not interrupting at inappropriate times. (60)*
2. Operations Analysis - Analyzing needs and product requirements to create a design. (60)*
3. Speaking - Talking to others to convey information effectively. (60)*
4. Writing - Communicating effectively in writing as appropriate for the needs of the audience. (60)*

5. Critical Thinking - Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, conclusions or approaches to problems. (56)

*(Note 1, 2, 3 and 4 have tied scores)

Despite the extensive research studies that underpin O*NET's methodology and conceptual framework, one of the criticisms against O*NET overall is that its depth of detail can be explained by a number of underlying constructs. Harvey & Wilson's (2010) research concluded that 4 factors could account for up to 72% of the common variance within the 173 individual measures. They argue that finding core factors to explain a large part of the variance, in most research circumstances, would be beneficial: the researchers would be able to refine their focus to a number of distinct constructs. However for O*NET, the 173 single items are required to be distinct constructs in themselves, and as such Harvey & Wilson's (2010) finding does raise concerns about O*NET's validity. For the purposes of this study, though, such critique need not undermine the conceptual distinctions made by O*NET with regard to tasks, skills, knowledge, and abilities, but it does seem prudent to compare the detail of O*NET's finding with other sources of information relevant to graphic design competencies.

2.6.2 National Occupational Standards.

In the UK, National Occupational Standards (NOS) aim to identify broad areas of competencies required for jobs within particular sectors: the design, or the audiovisual industry, for example. These standards are then used by training and educational organisation to inform specific qualification such a Higher Nationals (HN's), or National Vocational Qualifications (NVQ's). In relation to graphic design, two Sector Skill Councils (Creative & Cultural Skills and Creative Skillset) are responsible for the development of NOS in their respective sectors.

Creative and Cultural Skills' National Occupational Standards for Design (CCS, 2009), for example, is a suite of 46 units, each covering a particular topic ranging from "writing a design brief" through to "managing a budget within a design environment" (p.1). Within each unit is a section on performance criteria that in O*NET's taxonomy could appear to be classified as tasks (Table 5). In the unit Create Visual Designs (CCSkills, 2009b, p2) it states:

Table 5: CCS's Create Visual Designs performance criteria

You must be able to:

- P1 clarify type of visual design required
- P2 in discussion with relevant people, determine factors that may determine or affect visual design concepts including the design techniques
- P3 clarify the target user and audience
- P4 research other products, designs, images, artwork and other creative work that may inspire design ideas
- P5 obtain other relevant information that may influence the design
- P6 experiment with imaging techniques to create required visual design
- P7 explore a range of typographical and visual design elements to create designs
- P8 generate a range of visual design ideas that are technically feasible, meet the brief and provide creative solutions
- P9 evaluate initial design ideas and specifications against findings and discuss with the relevant people to select the final design concept
- P10 select the design technique and discuss with relevant people to ensure the final design meets the requirements of the brief
- P11 develop the design based on the final design concept using the selected design techniques
- P12 apply visual design principles to produce designs
- P13 save the visual design in an appropriate format that meets the technical parameters determined in discussion with relevant people
- P14 review the visual design to assess whether it meets the design brief and any technical specifications
- P15 discuss and confirm with relevant people, additional requirements or modifications and undertake any amendments

Each unit also identifies the knowledge required for a given topic (Table 6). In Create Visual Designs (CCSkills, 2009b, p.3) it states the following:

Table 6: CCS's Create Visual Designs knowledge required

You need to know and understand:

- K1 roles and responsibilities of other team members
- K2 the relationship between technical and creative aspects of a design
- K3 basic design principles of layout and composition
- K4 characteristics of imaging methods
- K5 visual design and communication principles
- K6 principles relating to the use of typography
- K7 copyright clearance procedures

Based upon these units organisation such as BTEC have developed Specialist qualifications for Graphic Design, underpinned by aspects of the NOS.

2.6.3 BTEC Specialist

BTEC's specialist qualification in Graphic Design is intended to provide the competencies required for preparation for work within graphic design. A number of levels are available (for example, QCF Level 3), and within levels a range of qualifications are possible (award, certificate or diploma). Each of these qualifications has its own requirements, but core to each is that the course provider will choose from four of the following units shown in Table 7.

Table 7: BTEC Specialist, Graphic Design unit options

- Typographic and Layout Design
- Typefaces and Letter Forms
- Words and Images in Graphic Design
- Graphic Image Making
- Information Graphics
- Website Design
- Graphics Media, Techniques and Technology
- Mixed Media Image-Making
- Design for Advertising
- Graphics for 3D Application
- Narrative Image Making
- Digital Story Telling
- Specialist Illustration Computer Applications
- Factual Writing in Art and Design

Each unit states the learning outcomes required and gives guidance on the content that could be expected. For example, with the unit Words and Images in Graphic Design, one of the three learning outcomes is to "Understand the relationship of words and images when developing ideas" (BTEC,2010, p1), and part of the content suggested is "Developing ideas: initial ideas e.g. roughs, storyboards, sketches, studies, layouts, preliminary designs; development e.g. reviews, refinements, analysis, evaluation; annotations e.g. notes, written pieces, descriptions; formats e.g. sketchbooks, notebooks, storyboards, worksheets, design sheets, onscreen" (p2).

2.6.4 BTEC Higher Nationals (Graphic Design) Level 4 & 5

Like BTEC Specialist qualification, BTEC Higher National in Graphic Design are also underpinned by NOS and suggested to reflect the needs of professional organisations and the graphic design industry. They are described as being "recognised internationally by employers as providing learners with the key concepts and practical skills for direct progression to, or within, employment" (BTEC, 2013, p1).

In addition to the selection of specialist units (such as Advanced Typographic Design, or Corporate Identity in Graphic Design), at level 5 (Table 8), the qualification is structured upon five core units:

Table 8: BTEC Higher Nationals (Graphic Design) Core Units

- | |
|---|
| 1 Visual Communication in Art and Design
2 Ideas Generation and Development in Art and Design
3 Contextual and Cultural Referencing in Art and Design
4 Professional Practice in Art and Design
5 Project Design, Implementation and Evaluation |
|---|

Each core unit has an abstract that outlines the broad aims of the unit, followed by sections on learning outcomes (Table 9) and suggestions for content and assessment criteria.

Table 9: BTEC Higher Nationals (Graphic Design) learning outcomes

BTEC HN's Core Units	Learning Outcomes
Visual Communication in Art and Design: learning outcome 1	Be able to communicate ideas and concepts by researching visual techniques
Visual Communication in Art and Design: learning outcome 2	Be able to select visual communication techniques to realise creative intentions
Visual Communication in Art and Design: learning outcome 3	Be able to produce work which demonstrates the use of visual communication
Visual Communication in Art and Design: learning outcome 4	Understand the potential for personal development through the application of new approaches to visual communication.
Ideas Generation and Development in Art and Design: learning outcome 1	Be able to exploit idea-generating techniques
Ideas Generation and Development in Art and Design: learning outcome 2	Be able to implement complex ideas visually
Ideas Generation and Development in Art and Design: learning outcome 3	Be able to express ideas using visual techniques
Ideas Generation and Development in Art and Design: learning outcome 4	Understand the cultural contexts in which ideas are visually transmitted.
Contextual and Cultural Referencing in Art and Design: learning outcome 1	Be able to undertake in-depth research
Contextual and Cultural Referencing in Art and Design: learning outcome 2	Understand influences on art and design activities and outcomes through the interpretation and analysis of information
Contextual and Cultural Referencing in Art and Design: learning outcome 3	Be able to assess, interpret and evaluate information
Contextual and Cultural Referencing in Art and Design: learning outcome 4	Be able to evaluate and present conclusions.
Professional Practice in Art and Design: learning outcome 1	Be able to place themselves and their work in the context of their selected discipline
Professional Practice in Art and Design: learning outcome 2	Understand their specialist area and the career opportunities available
Professional Practice in Art and Design: learning outcome 3	Be able to develop and present a professional portfolio in an appropriate format
Professional Practice in Art and Design: learning outcome 4	Understand how to promote themselves and their work professionally.
Project Design, Implementation and Evaluation: learning outcome 1	Be able to formulate a project
Project Design, Implementation and Evaluation: learning outcome 2	Be able to implement the project within agreed procedures and to specification
Project Design, Implementation and Evaluation: learning outcome 3	Be able to evaluate the project outcomes
Project Design, Implementation and Evaluation: learning outcome 4	Be able to present the project outcomes.

2.6.5 NASAD

NASAD (the National Association for Schools of Art and Design) is the organisation that accredits the Professional Baccalaureate Degree in Graphic Design in the US. It defines one of its key objectives at "To establish reasonable standards centered on the knowledge and skills necessary to develop academic and professional competence at various program levels" (NASAD, 2012, p.1). Alongside a number of broader competencies related to art and design, it describes six essential competencies that underpin their Graphic Design qualification. Each of these competencies is given a further description (p116), table 10.

Table 10: NASAD essential competencies for graphic design

Essential Competencies	Description
The ability to solve communication problems...	including the skills of problem identification, research and information gathering, analysis, generation of alternative solutions, prototyping and user testing, and evaluation of outcomes.
The ability to describe and respond to the audiences and contexts which communication solutions must address...	including recognition of the physical, cognitive, cultural, and social human factors that shape design decisions.
The ability to create and develop visual form in response to communication problems...	including an understanding of principles of visual organization/composition, information hierarchy, symbolic representation, typography, aesthetics, and the construction of meaningful images.
An understanding of tools and technology...	including their roles in the creation, reproduction, and distribution of visual messages. Relevant tools and technologies include, but are not limited to, drawing, offset printing, photography, and time-based and interactive media (film, video, computer multimedia).
An understanding of design history, theory, and criticism from a variety of perspectives...	including those of art history, linguistics, communication and information theory, technology, and the social and cultural use of design objects.
An understanding of basic business practices...	including the ability to organize design projects and to work productively as a member of teams.

2.6.6 Conclusion

As detailed above, each of these organisations considers they have identified the sort of knowledge, skills and abilities that a graphic designer requires (be this at various stages in their career development), and has undertaken their own specific research within graphic design to arrive at the titles and descriptions given. Whilst little of this research has undergone peer-reviewed in the manner of double-blind academic journals, many of these organisations state involvement from design practitioners, and input from industry, in their development, as a basis for validity.

An observation that can be made of the various competency frameworks for graphic design discussed is they lack a common language and structure to bring cohesion to the depth and breadth each offer. It is this lack of theoretical structure that makes the identification of graphic design competencies complex, complicated, and at times utterly confusing. Whilst no model will be perfect (McGrew, 2009), the benefit of a common theory with which future research can build, and past studies integrated and re-evaluated is, arguably, overdue.

One of the main aims of this section is to highlight the need for a more cohesive approach to graphic design competencies within the UK. Whilst it is beyond the scope of this work to identify a fully comprehensive model underpinned by empirical research, it is possible to formulate the general direction of travel, and begin to test some of the abilities suggested to important for graphic design with empirical findings from design creativity research.

2.7 Key Relationship: Skill and Creativity

As creativity is deemed an important indicator of competency as a graphic designer - some would argue for any designer (Cross, 1997) - then how are they related? What is the relationship between skill and creativity? Within the literature under study, at least four

types of relationship are present. These can be described as follows: a linear negative correlation; a linear positive correlation; a non-linear positive correlation, and a non-linear negative correlation.

2.7.1 A Linear Negative Correlation

Stacey, Eckert & Wiley (2002) in their research for textile design highlighted that in the professional experience of a number of the participants interviewed "...expert designers who have acquired technical knowledge are less creative" (p.5). Perhaps this is reflective of this specific domain, but it does state well the perception that there is a negative relationship between knowledge (technical in this case: possibly skills and abilities can be inferred) and creativity. Figure 2, gives a representation of how this relationship may look if the level of skills was plotted against the level of creativity.

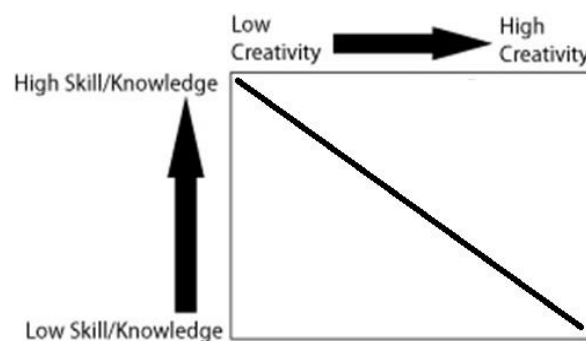


Figure 2: A negative linear relationships for skills/knowledge relative to creativity

2.7.2 A Linear Positive Correlation

In contrast, Amabile's componential model of creativity (Amabile, 1996; 1983) implies that highly skilled and knowledgeable designers are more likely to be creative and that the relationship of skills to creativity is a positive correlation. The more domain relevant skill and knowledge a designer achieves (combined with high levels of task motivation and

creativity relevant skills) the more likely they are to achieve a creative output in their domain (Figure 3). Such a position could also be advocated by Skillset: that the acquisition of domain skills and knowledge is positively correlated to creative output, and this is a linear relationship.

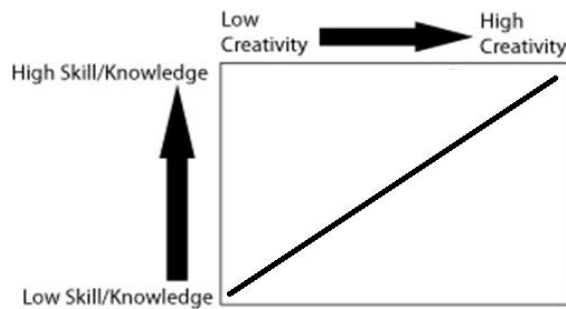


Figure 3: A positive linear relationships for skills/knowledge relative to creativity

2.7.3 A Non-Linear Positive Correlation

Haylock (1997) would agree that knowledge and creativity are positively correlated, but would consider the relationship to be non-linear. His research on mathematic knowledge and creativity in mathematics identified three groupings of student. First, those who had low knowledge in mathematics showed low creativity in mathematics. It would appear that a certain level of knowledge (possibly skill and ability) is required in mathematics in order to respond creatively to the mathematical tasks that Haylock set. Second, of those that did show high levels of mathematical knowledge, some were more creative than others. Haylock observed those who showed high knowledge/low creativity; with a third group that showed high knowledge/high creativity. Taken together, the relationship between these three groups of students suggests an entry threshold for creativity (Figure 4): overall a relatively high amount of mathematical knowledge is required to be creative in mathematics. But, this is not enough in isolation. Without the sort of creativity relevant skills, divergent thinking skills, or risk taking (as suggested in a number of creativity

theories) high levels of domain knowledge alone is insufficient for creative output within a domain.

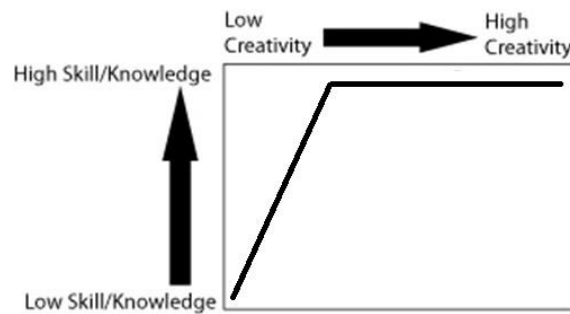


Figure 4: A positive non-linear relationships for skills/knowledge relative to creativity

2.7.4 A Non-Linear Negative Correlation

Where Haylock's findings suggest a person needs high levels of domain knowledge to be creative, Sternberg's (2006) work hypothesises it is possible for a person to have too much domain knowledge: that high levels of domain knowledge and skill can undermine creative potential (Figure 5). If Haylock's research suggests there is an entry point, then Sternberg's research suggests there is also an exit point: an exit threshold beyond which increased the level of domain knowledge actually restrict the novelty of ideas. How this may work is open to debate, but one possibility is that the considerable investment to acquiring high levels of domain knowledge may make individuals less open to new or radical perspectives than those with moderate levels of domain knowledge.

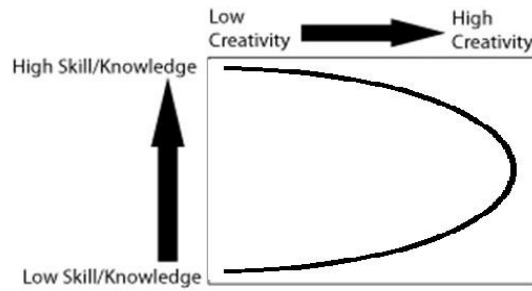


Figure 5: A non-linear negative relationships for skills/knowledge relative to creativity

2.7.5 Comparable Shapes of Correlations

As can be seen above with the attempt to represent these four different perspectives graphically; the main point is to illustrate the four differences of shape: be it representative of a linear negative correlation; a linear positive correlations; a non-linear positive correlation suggestive of Haylock's work; or a non-linear negative correlation suggestive of Sternberg's work. None are the same shape, and yet they each represent legitimate views on the relationship of skill/knowledge to creativity.

2.7.6 Influence of Domain on Skills/Creativity Relationship

Given the literature cited above, there is likely to be a relationship between skills/knowledge and creativity, but which one of these four is the most accurate?

To ask such a question is to imply that a general relationship may exist, yet, as discussed previously, that cannot be assumed: they may differ because of the domain. Haylock's work was for mathematics (and school children); Stacey & Eckert on textile design (and adult professionals); Sternberg and Amabile are theoretical models (and if they are to find validity it will be through domain-specific examples).

Thus, all of the perspective, caveats, conceptual frameworks and operational definitions

discussed previously can now be brought to bear upon the core relationship between skills, creativity and domain.

2.8 Core Relationship: Skills, Creativity and Domain

In the previous sections, each of the key terms in this study has been discussed: domain, skill and creativity.

In summary, the position taken in this study is to define domain at the level of a single occupation (graphic designer) and to control the sampling to include experts who identify themselves with this domain (but who may describe themselves as a graphic artist, visual communicator, etc.). Moreover, in this study, experts will be sampled from graphic designers working on a full-time basis.

An argument was presented that there are a number of complexities, and contradictions, regarding the usage of the term skills. But, for this study, research on expertise, particularly work in the area of vocational psychology, and within this job analysis, was able to clarify common language on terms and a hierarchical taxonomy with which to navigate this topic. Furthermore, a number of conceptual distinctions can be made between, skill, knowledge, ability, tasks and competency. In this study, the focus will be, for the most part, on abilities

Regarding definitions of creativity, perspectives can be taken that place emphasis on the person, process, product or press (environment). The stance taken in this study is to define creativity through the perspective of a creative output (product) and to assess this creative output through use of the Consensual Assessment Technique.

With these terms established, relationships between these key terms were explored: that is, between creativity and domain; domain and skill; skill and creativity.

In summary, the relationship between domain and creativity is polarised within the academic creativity research community: some advocate domain-specificity of creativity, other see it as a domain-general construct. Compromise theories have been offered, but they can be seen as not addressing the concerns. In the absence of agreement, an argument can be made to side with domain-specificity as being a reasonable basis to future-proof current research studies.

With respect to what skills, knowledge and abilities are important to graphic design, from a research perspective this remains obscure, yet, several organisations have findings that, in different ways, attempt to address this issue, and appear to show similarities and overlap. Thus, there appears to be a gap in the literature and a need for some form of broader analysis.

Notwithstanding the need for such work, with regard to the relationship between skill and creativity, at least four types of relationship can be found in the literature, and are suggestive of contradictory findings. Some suggest the relationship is negative, others positive, some linear, others non-linear. Whether such contradictions are due to the domain being studied remains a consideration for future research.

Finally, returning to the value and impact of this topic, our understanding of the relationship between skill, creativity and domain have impacts that are theoretical, educational and economic.

2.8.1 Skills for Creativity in Graphic Design

Bringing all these influences to bear on the core relationship between skills, creativity and domain, at the outset, the broad research direction was stated as "identifying those skills that a domain considers crucial for competent performance, and testing their relationship (if any) to creative output within a domain". With a clearer appreciation of such terms and relationships, the core research question is restated as follows:

*To what extent are domain competencies (as defined by O*NET's version 16.0 job analysis database for graphic design) correlated with creative output in graphic design (as measured by the Consensual Assessment Technique).*

The specific focus of this question on creativity, competencies and graphic design appears not to have been asked by other researchers or organisations. The closest possible match, to date, has been the work of Mackie (2005).

Multiple Intelligences and Graphic Design Ability in Five North Carolina Community Colleges was Mackie's Ed.D thesis, and a research study based on 113 first-year graphic design students. In this work, he identified that linguistic intelligence might be more important to graphic design ability than visual-spatial intelligence. Whilst the intent of the study was to explore the theory of Multiple Intelligences, as proposed by Howard Gardner (1999), within the context of graphic design, such constructs as visual-spatial intelligence could be framed as abilities (some researchers use the term visual-spatial abilities). Equally, graphic design creativity was not directly measured, but the assessment of graphic design ability was based on a portfolio of "several commercially appropriate, graphic-related works, and faculty evaluated the project portfolio using their regular evaluation procedures based on standards established in their programs" (Mackie, 2005, p.59). It is likely, given the emphasis on creativity in other sections of his thesis, that

creativity was an implicit, if not explicit, criteria for the faculty undertaking the portfolio assessment.

Notwithstanding these distinctions, the implications of Mackie's findings are important and somewhat counterintuitive to many in art and design: indeed Mackie himself suggests: "The ability to recognize and understand shape, form, and space is essential to the visual arts, particularly graphic design" (p.85). Yet, the implications of such research suggest that greater focus should be placed on the development of reading and writing abilities, and ironically, less emphasis on drawing ability. If this is the case, this would suggest a significant re-evaluation of educational traditions within graphic design curriculum, and across secondary, further and higher education.

However, is one study enough on which to base a substantial change in graphic design education? Very little research has been conducted on graphic design creativity and specific cognitive abilities. Equally, researchers understanding of how to measure cognitive abilities like mental visualisation has become more nuanced and, as will be explored later, the measures Mackie used in his 2005 study are only part of a new understanding of what may constitute a holistic view of cognitive visualisation (it can be argued that the test Mackie used are now known to be biased towards more scientific and engineering backgrounds than design and artistic professions).

Thus, whilst Mackie's findings are important, they also stand in isolation, and direct measures of a key construct under investigation: graphic design creativity, has yet to be explored within this context. As a result, no published research has been identified that has applied the CAT to graphic design as a means to test the relationship between creativity and domain competencies. Indeed, whether this relationship is linear, non-linear,

or non-existence has yet to be explored.

Such a research focus, however, given the literature review findings discussion so far, highlights two specific obstacles that have implications for experimental research design and method in order to move forward

Firstly, identifying competencies in graphic design is not straightforward. Even with O*NET's work as a basis for definition, it is not obvious how to operationalize many of the factors suggested as important skills, knowledge and abilities in relation to graphic design for research purposes. In part this is because O*NET's descriptions, oddly enough, lack detail; conversely, those organisation that have detailed findings on graphic design competencies appear to lack a suitable conceptual framework on which to distinguish between skills, knowledge and abilities. The result is a level of complexity that suggests some form of analysis and coding of these respective works is required, and would be beneficial. This is particularly important as in order to test a given correlation it is likely that not all domain competencies that O*NET has identified can be tested at once, and may need to be prioritised. The need for selection is pragmatic in terms of the time and resources available to a single researcher governed by the customs of a PhD (such as the need for sole authorship and undertaking all fieldwork). It is also the case that a battery of tests would be extensive (especially if more than one test is required for each construct) and participant fatigue needs to be taken into account purely on ethical grounds, if not for research design reasons and attrition rates. Furthermore, whilst some competencies may be important, without a suitably valid and reliable measure, would lack research credibility as a useful construct to explore. For all these reasons, a more specific study of the interactions between O*NET and other graphic design competency models is required.

Secondly, and related to above, as creativity is the dependent measure for this research question, it is crucial that its operationalization should find a sufficient level of acceptance within the creativity research community and graphic design practice. Without such acceptance, regardless of the outcome, the findings of this study would have little value. In this respect, the CAT certainly has a longstanding pedigree as a suitable measure of domain creativity. Unfortunately, at present, its specific application to graphic design (possibly design more broadly) is not well understood. Whilst it is fair to argue for the CAT as a measure of creative output, exactly what constitutes a valid and reliable graphic design task or design brief to be given to participants appears to be unknown.

For these reasons, with each of these obstacles identified, two studies were undertaken; the details and results of which form the next and proceeding chapter of the thesis.

Chapter 3, Study 1: A Systematic Review of the CAT in Design Research

3.1 Introduction to Study 1

In the previous sections, creativity was defined as a creative output to be assessed using the CAT; domain at the level of a single occupation domain (graphic design in this instance). Based on these definitions to what extent has CAT being used in design research, and specifically graphic design creativity research?

As discussed in the defining creativity section (2.3), for the past thirty plus years Amabile's (1982) Consensual Assessment Technique (CAT) has been used as a reliable and valid measure of creativity. It has been described as the "gold standard" of creativity assessment (Baer & McKool, 2009); been extensively used within creativity research, and is seen as the most popular method of assessing creative outputs (Kaufman, Plucker & Baer, 2008).

Its discussion within scholarly research has continued to grow year by year. However, since 1996, a systematic review of the CAT has not been undertaken within creativity research, and, within design research publications appears to have never taken place.

At the time of Study 1, despite this historic place within creativity research, the use of the CAT as a measure of creativity within design research appears to be relatively small. For example, within the published proceeding of invited papers for Design Creativity 2010, only one citation was given for the CAT (Collado-Ruiz & Ostad-Ahmad-Ghorabi, 2010b). The main aim of Study 1 was to consider whether this was an accurate reflection of the use of the CAT in design research. A systematic review of the CAT was undertaken to

create a comprehensive database of CAT citations with which to:

- Identify key journals publishing CAT studies
- Identify key authors publishing CAT studies
- Identify the use of the CAT within a range of design journals
- Identify the use of the CAT in relation to graphic design

3.2 Systematic Review

A systematic literature review is considered a substantial form of research in its own right, and quite different to the traditional literature review found in academic publications.

Indeed, Gough, Oliver & Thomas state (2012, p.4) that:

So influential has the use of research through systematic reviews become that their development can be considered to be one of the turning points in the history of science: "This careful analysis of information has revealed huge gaps in our knowledge. It has exposed that so-called 'best practices' were sometimes murderously flawed; and by doing nothing more than sifting methodically through pre-existing data it has saved more lives than you could possibly imagine" (quoting Ben Goldacre's comments on systematic reviews for BBC Radio 4's Moment of Genius)

Whilst Goldacre's view is particularly relevant within a medical research context (for example, such as a Cochrane Review), other researchers, such as those at University College London's Evidence for Policy and Practice Information (EPPI), have established the value of systematic reviews for social science purposes. While a systematic review of the CAT is not as emotive a topic as challenging areas of medical research or social science, the common ground to establish is the emphasis on method that distinguishes

this approach from more traditional reviews: a key feature being a systematic method, with the aspiration to detail a high level of transparency in respect to the inclusion of some studies over others.

3.3 Method

The CAT reference database was built using Reference Manager and Zotero to import citations from a number of databases, namely: PsycINFO, ISI Web of Knowledge and Google Scholar. Details and rationale are provided below. Key journals were defined as those with more than five citations for CAT within the last 31 years (between 1980 and 2011); key authors were defined as first authors with more than 10 citations for CAT within the last 31 years.

3.3.1 Why use Google Scholar Rather than Other Databases?

Google Scholar has its advocate and critics. The reason for its inclusion in this study was as a direct result of the low number of citations to be found in more esteemed databases such as PsycINFO and Web of Knowledge for the consensual assessment technique. Given the aims of this study to undertake a systematic and comprehensive review, such omissions were major concerns. Without discounting the inaccuracies and limited data available via Google Scholar, its database offered the most comprehensive list of references related to the Consensual Assessment Technique. Given this, the strategy used was to gather all of the related sources available via Google Scholar and then check for duplications within PsycINFO and Web of Knowledge. The search criteria used were the term "consensual assessment technique", with the following restrictions: articles excluding patents, anytime, at least summaries, English-only documents. The result was 737 citations that matched these criteria on the 25th of January 2012.

3.3.2 Downloading Multiple Citations from Google Scholar

At the time, Google Scholar only allowed downloads of one reference at a time, but with the application of a Zotero plugin for Mozilla Firefox multiple downloads were feasible. Unfortunately, this does have restrictions, and it was not possible to download all 737 citations from Google Scholar in one go. The option was to download a set number of citations at a time, with the need to refresh Google Scholar each time after an error message appeared. Doing so, however, led to counting inaccuracies between Google Scholar, Zotero and Reference Manager; with the possibility that either Google scholar was not correctly counting the citations (as has been suggested by other researchers), or something was getting lost in the process of exporting. The approach taken to account for this was to download a decade of citations, or any number of years up to a limit of 100 citations at one time, and double check for the number of citations counted into Zotero and then exported to Reference Manager. Using this method, Zotero counted more citations than Google Scholar had suggested for 2004-2005 and 2008. Given that these figures were higher the decision was taken to accept the Zotero figures as the more accurate. Discounting for papers in 2012, the total number of citation exported to Reference Manager was 745.

3.3.3 Cleaning the Reference Manager Database

A duplicate search of these 745 citations, highlighted 14 cases of duplication (i.e. seven pairs of matching papers). The criteria set for duplication was author and title. Six citations were exact duplicates. Thus, half of these duplicates were deleted: Hickey et al., 2006; Vrani-ç, 2005; and an undecipherable name (2005). The remaining 8 papers were retained for the following reasons: Sternberg (1995) and Sternberg et al. (2005) shared the same title in this database but were clearly different papers; Rhee (2006) and Rhee (2007) shared the same title, but one was a conference proceeding and the other a journal paper; Mathing et al. (2004), and Sanden et al. (2006) shared the same title, but

one was a book chapter and the other a journal paper; Christensen (2006) and Christensen (2006) shared the same title and date, but one was a thesis and the other a journal paper; leaving a total of 742 citations in the database.

Despite setting the search criteria for papers written in English, a further 19 citations were identified that the titles suggested were written in a different language. These papers were deleted from the database. Furthermore, 10 papers had data that was undecipherable in the form of symbols and were also deleted from the database; leaving a total of 713 references in the database.

3.3.4 PsycINFO

In contrast to the several hundred citations for CAT within Google Scholar, PsycINFO returned 45 citations for the Consensual Assessment Technique. Within this list, none of the design journals in this study were cited. After checking for duplicates between the Google Scholar database and PsycINFO (of which there were 39), this led to the inclusion of the following 6 references: Conti et al. (1996); Baumgarten (1997); Mannarelli (2000); Liu & Shi (2007); Batey & Furnham (2009); Tan (2009); leaving a total of 719 citations in the database.

3.3.5 Web of Knowledge

Similarly to PsycINFO, Web of Knowledge returned 54 citations for the Consensual Assessment Technique. After checking for duplicates within the updated database (of which there were 51), this led to the inclusion of the following three references: Corko & Vranic (2007); Hennessey & Amabile (2010); Kaufman (2010); leaving a total of 722 citations in the database.

3.3.6 Design Research Journals

Nineteen journals were chosen to represent design research for this review. They were as follows: *Artifact*; *CoDesign*; *Design Issues*; *The Design Journal*; *Design Philosophy Papers*; *Design Studies*; *Form*; *International Journal of Arts and Technology*; *International Journal of Art & Design Education*; *International Journal of Design*; *International Journal of Design Sciences & Technology*; *International Journal of Technology and Design Education*; *Journal of Design History*; *Journal of Design Research*; *Journal of Engineering Design*; *Leonardo*; *Scientometrics*; *Social Studies of Science*; *Technoetic Arts*.

3.4 Results

3.4.1 CAT Citations

Based upon the database described above, Figure 6 shows the growth of CAT citations relative to design related journals from 1980 to 2011.

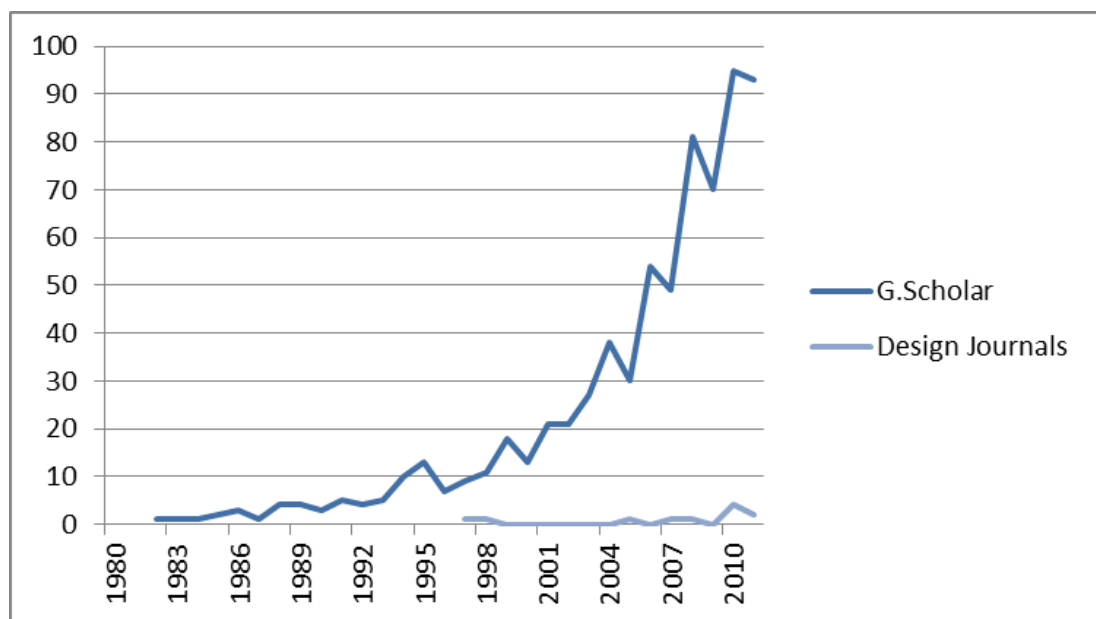


Figure 6: Number of CAT citation (per year) in design journals relative to other journals

3.4.2 CAT Citations Within Design Journals

In searching for a specific reference to the Consensual Assessment Technique within design journals, a total of 11 papers were identified in the following journals (in date

order):

- Cross, N. (1997). Descriptive models of creative design: application to an example. *Design Studies*
- Verstijnen, I. M., Van Leeuwen, C., Goldschmidt, G., Hamel, R., and Hennessey, J. M. (1998). Sketching and creative discovery. *Design Studies*
- Christiaans, H. and Venselaar, K. (2005). Creativity in design engineering and the role of knowledge: Modelling the expert. *International Journal of Technology and Design Education*
- Jeffries, K. K. (2007a). Diagnosing the creativity of designers: individual feedback within mass higher education. *Design Studies*
- Kokotovich, V. (2008) Problem analysis and thinking tools: an empirical study of non-hierarchical mind mapping. *Design Studies*
- Collado-Ruiz, D. and Ostad-Ahmad-Ghorabi, H. (2010a). Influence of environmental information on creativity. *Design Studies*.
- Cropley, D. and Cropley, A. (2010) Recognizing and fostering creativity in technological design education. *International Journal of Technology and Design Education*
- Pektas, S. T. (2010) Effects of cognitive styles on 2D drafting and design performance in digital media. *International Journal of Technology and Design Education*
- Howard, T. J., Culley, S., and Dekoninck, E. A. (2011). Reuse of ideas and concepts for creative stimuli in engineering design. *Journal of Engineering Design*
- Jeffries, K. K. (2011). Skills for creativity in games design. *Design Studies*.
- Lau, K. W. (2011). The difficulties of assessing design students' creativity: a critical review on various approaches for design education. *Journal of Design Research*

Of these 11 papers, two operationalize the CAT within their studies and state this specifically (Christiaans & Venselaar, 2005; Pektas, 2010). Three papers use judges to evaluate creative outputs, but these are not directly related to the CAT in terms of procedures (Verstijnen et al., 1998; Kokotovich, 2008; Collado-Ruiz & Ostad-Ahmad-Ghorabi, 2010a). The remaining papers up to 2011 make reference to Amabile's 1982 or a CAT work to support a point within their papers, but CAT was not part of the study (Cross, 1997; Jeffries, 2007; Cropley & Cropley, 2010; Jeffries, 2011; Howard, Culley &

Dekoninck, 2011; Lau, 2011).

3.4.3 Key Journals Publishing CAT Studies

From the 722 database, three journals stand as key contributors to the debate on the Consensual Assessment Technique (number of CAT citations found in the database in brackets). These are:

- Creativity Research Journal (55, CAT citations)
- The Journal of Creative Behavior (25)
- Psychology of Aesthetics, Creativity, and the Arts (18)

The caveat to this claim is that 207 of the reference in the database were without data on journal, books, chapter or thesis. However, 515 citations remained with these details, and it was decided that this number of citations formed a reasonable basis on which to proceed. Those with 5 or more reference to the CAT were as follows:

- Journal of Personality and Social Psychology (11)
- Roeper Review (11)
- Creativity and Innovation Management (11)
- Personality and Individual Differences (8)
- Thinking Skills and Creativity (7)
- Design Studies (7)
- Journal of Research in Personality (6)

The following books and dissertations were also identified by the database:

- Dissertation Abstracts International (9) (dissertations)
- The Cambridge Handbook of Creativity (7) (book)
- Creativity: When East meets West (5) (book)

3.4.4 Key Authors Publishing CAT Studies

From the database of CAT citations six first authors account for 91 of the papers as

follows (number of citations in brackets):

- Kaufman, J. C (22)
- Baer, J (17)
- Dollinger, S. J (15)
- Amabile, T. M. (13)
- Hennessey, B.A (13)
- Plucker, J. A (11)

3.4.5 CAT Studies Related to Graphic Design

As the CAT database described above had limited abstracts details, initial searches for graphic design showed no results. Subsequent follow-up searches using Google Scholar identified 24 papers in the CAT database that made reference to both graphic design and the Consensual Assessment Technique. The vast majority of studies were not related to graphic design directly: for example, a paper may have mentioned graphic design software, but neither the area under study nor the types of participants were related to graphic design. The exception being, two papers appeared to make use of both CAT and graphic design: Silvia et al. (2008) and Dineen & Niu (2008).

3.5 Discussion

Reflecting on the use of CAT in design research, one of the most striking findings of this systematic review was just how few CAT studies had been published in design journals. The reasons for this are likely to be complex and interrelated, but a number of distinct themes can be identified and are discussed below: creative process versus creative output; criterion versus normative assessment; concerns of the methodological stance of the CAT, concerns on the specific method of applying the CAT.

3.5.1 Creative Process Versus Creative Product

One theme can be seen in the classic dichotomy between the creative process and creative product. In an earlier section of this thesis, the definition of creativity being based on outputs that had the dual quality of originality and usefulness was highlighted. This type of definition is frequently expressed in academic journals with the inference this is a common consensus amongst scholars of creativity (Kaufman & Baer, 2012). Equally important, however, is the controversy surrounding how researchers' define core characteristics like originality and the type of evidence required to assess creativity (Runco, 1999). To help resolve this, in recent years, a number of works now categorise creativity in more precise terms, such as Big C creativity, Pro-C Creativity, Little C creativity, Mini C creativity (Beghetto & Kaufman, 2007; Kaufman & Beghetto, 2009), and have different expectations associated with each of these terms regarding originality and usefulness, and the sort of study population they can be applied to. For example, kindergarten pupils, students in higher education, professional practitioners within a field, or those who have made an eminent and outstanding contribution to a domain.

Furthermore, in section 2.4.3, the focus on creative output was presented as not the only means to define creativity; other categories of research explore the creative process, the creative person, or the creative environment (Isaksen et al., 1993). The argument can be made (for some creativity researchers this argument is generally accepted) that, to validate findings in these three other areas, inference needs to be shown with the creative output (Kaufman & Baer, 2002). By doing so, this places the primacy of creative output at the heart of creativity assessment. As mentioned above, such a decision is controversial. It may also partly explain the low use of the CAT in design research, as highlighted by the findings of this systematic review.

The CAT is a measure of creativity that is firmly placed within notions of creativity as an

output. With a degree of contrast (as a broad generalisation) much focus within design research has been given to "understanding designing as a process" (p.1, Nagai & Gero, 2012).

Such broad distinctions are not to be taken as a sign of polarity: while offering theoretical value, each of the 4 P's of creativity research (product, process, person, press) clearly interact and influence each other in practical terms. Indeed, the number of methods with which to measure creativity are considerable and varied (Batey & Furnham, 2006): from protocol studies, self-report measures, divergent thinking tests, to creativity assessment by domain experts, and each method has its strengths and limitations.

Notwithstanding the value of alternative methods to researching creativity, perhaps an emphasis on the process of design creativity has overshadowed methodologies that relate more to the creative output? The rationale for asking this question is based on the small number of CAT studies in design journals relative to creativity journals and the number of CAT citations within the wider community of scholars.

The counter argument is that those studies within design journals that cite the CAT often directly address this issue of product relative to process; notably Cross (1997) and Christiaans & Venselaar (2005), but in both cases, this could well be traced back to Christiaans earlier research in 1992. Thus, one could argue that the value of creative output methodologies, and by extension, the CAT, has had eminent supporters within design research for some time.

However, whilst the foundations have been in place, when it comes to the CAT as an operational measure of creativity, on the systematic basis of this review, few in design

have built upon this. Of those studies that cited CAT in design journals, from the 11 papers identified, only two directly made use of the CAT. As will be discussed later, neither study was concerned with graphic design. In both studies, the reliability of the CAT was, for the most part, above the standard 0.7 levels required for inter-rater reliability (ranging from 0.66 to 0.81; 0.81 to 0.93).

Whilst these results show mostly favourable levels of consensus, the number of expert judges varied from 10 graduate industrial design students to 3 design academics. Selection of judges in terms of the level of domain expertise, and how many judges should be used in a study is a point of debate. Issues around the use of novices relative to expert judges have for and against arguments; equally, even within psychologically focused creativity journals, the number of judges has varied in CAT studies throughout the years. These are issues of interpretation of CAT method, to be discussed shortly, but interwoven with issues of method are issues of methodology, and here too, the CAT could also have been at odds with the milieu of design creativity research.

3.5.2 Method and Methodological Concerns with a Social Consensus View of Creativity

Lau's 2011 paper, *The difficulties of assessing design students' creativity: a critical review on various approaches for design education* for the Journal of Design Research, is of particular interest at this point. Clearly, this is a paper that undertakes a detailed review of methods to assess creativity and relate them to design. It covers a wide range of works that fall within self-report measures, divergent thinking tests, creative problem solving, and specifically discusses the creative output, or end product, as a measure of creativity within design.

The connection to CAT is established indirectly through citing Hennessey's work (1994),

however, given the detailed focus within the paper on TTCT, Creative Problem Solving, and other notable methods, it is interesting to consider why there is no direct discussion of the CAT. Indeed, the issue of subjective judgments and assessment criteria are fundamentally re-framed by use of CAT methodology. As mentioned earlier in this thesis, by emphasizing the subjective element in creativity assessment the CAT is at the opposite end of the assessment spectrum to TTCT, and is a method that is premised on the view that the judgments required to assess creativity "...can ultimately only be subjective" (Amabile, 1982, p. 1001), and that objective criterion does not currently exist (and may never exist) on which to assess creativity.

As a broad generalisation, much of the design research community has roots in industrial design and engineering, and for some, the search for algorithms, mathematical and computational models is the guiding focus of their research. Coupled with values of scientific objectivity, reductionism and an epistemological worldview (Creswell, 2013) likely to be aligned with a positivist paradigm, it is highly possible that a social consensus view of creativity is not an immediate cultural fit within this context. In this respect, does the philosophical stance of the CAT devalue the method from the perspective of design research?

If it does, then whilst this theoretical disconnect is understandable, it also seems at odds with aspects of design education, practice, and research. For example, the consensus of domain judges (often using their subjective opinion and expertise) remains a prevalent method for assessment in design education and professional design awards (of which creativity is either an implicit or explicit expectation). Equally, the sampling of eminent designers for design research (even when the focus is on understanding their creative process) has been premised on creative achievement among peers: Donald W.

MacKinnon's creativity research on architects (1962); Bryan Lawson's (1994) studies of successful architects; Nigel Cross's (2001) studies of outstanding designers, or Robin Roy (1993) studies of product designers, like James Dyson, are all seminal examples.

It could also be the situation that the design research communities concern about the CAT may be less about methodological consideration and more about the method itself; particularly the CAT's validity as a measure of design creativity.

As highlighted by the exponential growth in citations (Fig. 6), the CAT has been widely used in many disciplines and settings outside of design; at all educational levels (from kindergarten to higher education), and across numerous professions, both those traditionally associated with the Creative Industries and sciences, and in more obscure domains, such as the military. In this regard, one of the biggest strengths of the CAT is its (seemingly) simple method, and its adaptability to a wide variety of domains. However, it is precisely this deceptive simplicity and adaptability which may also be a barrier to its use within the design research community.

Problematically, the way Amabile's broad guidelines on the method are interpreted and implemented in practice by different researchers shows wide variation. In spite of guidance by Kaufman et al. who dedicated a complete chapter of their book *Essentials of Creativity Assessment* to the CAT, as did Hennessey, Amabile and Mueller for the *Encyclopaedia of Creativity*, substantial variation across research studies can be found in relation to CAT method.

For example, within the CAT literature, some researchers have created instructions that directly ask judges to discount technical execution and aesthetic appeal from their

creativity rating (Baer, 1993). Other researchers (the majority of CAT studies) have not done so with little adverse impact on inter-rater reliability: some ask judges to rate creativity alongside technical execution and aesthetic appeal (Amabile, 1982; Christiaans & Venselaar, 2005; Valgeirsdottir, Onarheim & Gabrielsen, 2015), some only do this the first time they undertake a new CAT task (Hennessey, 1994; Kaufman, Plucker & Baer, 2008); some do not. Thus, the important issue of technical execution and aesthetic appeal in relation to creativity, ranges from explicit, through to implied, implied only once, or not mentioned at all.

It is possible, such a degree of variation in details of the method could have been off-putting to the design creativity research community, and explain the low uptake in the CAT.

3.5.3 Criterion Versus Normative Assessment

A further reason for the CAT's lack of usage in design creativity assessment could be the fact that the CAT has traditionally being used for social psychology and group research rather than as an individual indicator or predictor of creativity. As has been discussed, the CAT is quite different to more individual scoring assessment measures like the TTCT. CAT ratings are based on the normative assessment of creative output within a defined group (i.e. the particular participants in a study), rather than individual scores standardised through defined criteria. In the case of TTCT, where standardised scores are available, standardisation is based on data gathered over decades of TTCT usage and responses. Individual responses are compared to a body of existing responses, with the comparison being made by trained TTCT assessors familiar with the characteristics of the population data. One of the purposes of matching and pairing response allows assessors to consider the statistical frequency of a response based on how often the characteristics of that

response appear. In this way, it is possible to gather both where an individual response sits within a general population, but also how it compares longitudinally across time. In this regard, it is possible to compare TTCT scores and responses from the 1960's to those of the 2000's. This comparison is exactly what Kim did in her 2011 study, which caught the imagination of global media through such eye-catching cover stories as *The Creativity Crisis* (Bronson & Merryman, 2010) and *Your imagination is not playing tricks on you - children really are becoming less creative, study shows* (Daily Mail Reporter, 2011).

Such details of method are re-emphasised to highlight this is fundamentally different to the CAT. The initial reason for the CAT was to distinguish higher creativity from lower creativity within a group as a whole. Such a group would likely have been stratified for experimental research conditions: a set of participants may be placed in one condition (for example, on the influence of financial reward on creativity), another set of participants placed within a controlled condition; but the creative output of each (experimental and control condition) would be pooled as one group, and creativity ratings gathered from judges would evaluate the output from both groups without knowing which output was from which group. It is the relatively higher or lower creativity within those groups, rather than the creativity of the individuals who form them that is being assessed.

From a design research perspective, whether this is an asset or a limitation of the method is very dependent on the nature of the study being undertaken. It is quite possible that all the participants within a CAT study could be highly creative, or conversely all low, or all mediocre relative to the expectations of a domain. The CAT, in its present method, has little way of knowing if a sample is biased toward one extreme or another other. What it requires of judges is to make distinctions within a specific group regarding those outputs that are more creative than others, relative to the group as a whole. In studies on the

social psychology of creativity, this normative assessment is an asset.

For some design studies, however, these original constraints placed on the CAT as a normative measure of creative output, may have become stretched in the pursuit of aspirations to understanding more about the creative process of designers. For example, Casakin & Kreiler (2011) study on *The cognitive profile of creativity in design* applied the CAT to 52 students of architecture. Each student was tasked to create a "creative design for a small museum" (p.163), and these works were assessed by three judges. Inter-rater agreement was acceptable, .73, and upon this, participants were grouped in either being classed as "highly creative" (p. 166) or "those low in creativity" (p. 166) based on whether participants had scored above the mean creativity assessment, or below it.

For the reasons mentioned above, using a CAT to infer qualities of individual creative ability is problematic. Certainly, some creative outputs were assessed as more creative relative to the others, but this does not validate that a participant is "highly creative" (p.166), or that their cognitive processes are reflective of a high level of creativity or a lack of it.

A further point could be argued that setting the threshold for those considered to be highly creative as scores above the mean would be out of sorts with other measures to testing individual creative ability: TTCT test scores can be understood relative to a population mean, standard deviation, and interquartile range. Setting aside the ethical considerations of identifying the *highly* from *those low in* creativity, the operational definition of these terms is more likely to be found in the upper and bottom interquartile range of scores.

As such, have some design studies extended the method and technique of the CAT

beyond the CAT's theoretical framework; and, unfortunately, without the evidence to suggest that such extensions are valid, has this been problematic for the acceptance of the CAT by design researchers? It is difficult to affirm with confidence. Casakin and Kreiler (2011) is an example of a design study not published in one of the design journals selected for this review but one with a broader focus: *Thinking Skills and Creativity*. To what degree design researchers have formed a view on the CAT influenced by creativity research publications outside of established design journals is speculative.

What can be said is that of those few CAT studies published in design journals, Christiaans & Venselaar (2005) for example, have stressed the normative basis of the creativity scores through a research design that takes the extreme high and extreme low creativity scores within a group as the basis for their analysis. As a result, in their study two, 55 students took part, randomly sampled from the first-year population of 240 industrial design engineering students. With this sample, based on 10 judges independently rating a number of design projects, CAT inter-rater reliability was excellent, ranging from .89 to .93. Taking these scores, extreme high and low creativity scores were established, and 18 students were included as part of the qualitative analysis for the study.

Aside from the point mentioned on technical executions and aesthetic appeal, and a researchers' method of distinguishing high from low levels of creativity using CAT scores, other issues of method are the details of the task/design brief set; the amount of time given to participants to complete the task/brief; the domain identity of the judges; the number of judges. All these considerations could have an impact on CAT validity in relation to design creativity and be the basis for its low usage in design research.

Each of the issues above highlights that consistency of CAT method and transparency of reporting within journal papers has yet to achieve a level of stability that is accepted within the international creativity research community.

For the moment, such considerations will have to be put to one side, but these will be revisited in the broad discussion towards the end of this thesis, and also underpin several of the studies to come. The key finding to take from above is that there are a very limited number of CAT studies within design journals. The reasons for this are genuinely unclear. Is this a gap in the literature symptomatic of a preference for the creative process over the creative product; methodological bias toward social consensus within design research; a lack of clarity on CAT method, or, based on the type of design journals searched, while comprehensive, have some publications been overlooked in this study? A further explanation is that prior to this study, no other researcher had undertaken a systematic review of the CAT in design, and the data presented here was not available until the 2012 International Conference on Design Creativity.

3.5.4 Use of CAT in Graphic Design

As highlighted in the results section, the CAT database had limited abstracts details, and initial searches for graphic design showed no results. Subsequent follow-up searches using Google Scholar identified 24 papers in the CAT database that made reference to both graphic design and the Consensual Assessment Technique. The majority of studies were not relevant to the use of the CAT in graphic design with the exception of Silvia et al. (2008) and Dineen & Niu (2008).

Silvia et al. (2008) undertook a study to validate a new method of scoring divergent thinking tasks. Part of the study had participants who majored in arts subjects (accounting

for 9% of participants in the study), of which some of these participants majored in graphic design. Whilst the new technique involved subject rating, and it built the case for the validity of subject rating by citing CAT studies, the method used was not the CAT protocol. Indeed, they acknowledged the importance of expert domain judges for "studies of real creative product" (p. 70) but argued that this need not be the case for divergent thinking tasks assessment.

In contrast, Dineen and Niu's work (2008) utilised the CAT method, and involved participants in their 2nd year of study in graphic design at a Chinese art and design higher education institution. The study explored the respective merits of UK creative pedagogy relative to traditional Chinese pedagogy and ran different styles of creativity workshops based on a control group/experimental group research design. It is arguable, however, how closely aligned to graphic design the final outputs from either group were (Figure 7), shows an illustration of artworks presented to judges for CAT assessment).



Figure 7: Images of artworks selected for Dineen and Niu (2008)

Distinctions that can be made between graphic design and illustration design do not

undermine the value of Dineen and Niu's research (2008); for the purposes of this review it highlighted that at the time of this study, there was at best only one study that had applied the CAT to graphic designers, and it is debatable if the creative outputs rated in this study were graphic design or illustration design. How important the distinction between these two domains is partly answered by the relationship between skill and domain and the definition of domain already described for this thesis. As was presented in the earlier sections, the position taken in this study is to define domain at the level of a single occupation (graphic designer), and to control the sampling to include experts who identify themselves with this domain (but who may describe themselves as a graphic artist, visual communicator, etc.). It is interesting to note that O*NET job analysis makes a distinction between the role of graphic designer and that of illustrator. In O*NET's taxonomy, Illustrators are titled as *Fine Artists. Including Painters, Sculptors, and Illustrators*.

3.6 Conclusion

For the past thirty plus years, Amabile's (1982) Consensual Assessment Technique (CAT) has been used as a reliable and valid measure of creativity. It has been described as the "gold standard" of creativity assessment (Baer & McKool, 2009); been extensively used within creativity research, and is seen as the most popular method of assessing creative outputs (Kaufman, Plucker & Baer, 2008).

This systematic literature review of CAT citations highlighted the small number of studies in design journals that have used the CAT as a research method, and that within the domain of graphic design it is arguable if any CAT studies have taken place. Prior to this study, no other researcher had undertaken a systematic review of the CAT in design, and the data presented here was first discussed at the 2012 International Conference on

Design Creativity.

The reasons for the low number of CAT studies in design journals is genuinely unclear. Is this a gap in the literature symptomatic of a preference for the creative process over the creative product; methodological bias toward social consensus within design research; a lack of clarity on CAT method, or, have some publications been overlooked based on the criteria for inclusion in this study?

Of those studies that operationalise creativity through the CAT, at the time of this review, it appeared that only one study had been undertaken that was directly related to graphic design. It can be argued that the focus of this study was more aligned to illustrative design, and because of this, no precedent had been established for a task to measure graphic design creativity using the CAT.

Especially given debates on task selection and creativity assessment (Lubart & Guignard, 2004; Reiter-Palmon et al., 2009), and the importance of task selection as the means to establish the CAT as a measure of "real" creative output, the need to identify an appropriate task to measure graphic design creativity was not obvious. This review highlighted that such research needed to be undertaken, and resolving this gap in the creativity research literature became the focus of chapter 5, and study three, in this thesis.

Chapter 4, Study 2: A Conceptual Framework for Prioritising Graphic Design Competencies.

4.1 Introduction to Study 2

Returning to the core relationship between skills, creativity and domain, one of the obstacles identified in Chapter 2 was how to operationalise, for research purposes, the factors O*NET suggested as important skills, knowledge and abilities in relation to graphic design. The challenge presented was in part a mismatch between the level of detail available in O*NET's findings with that of other organisation more directly associated with graphic design training and education (BTEC, NASAD). The suggestion was to undertake a mapping study on the interactions between O*NET and other graphic design competency models. Specifically, to explore how existing conceptual frameworks that underpin O*NET may be used to prioritise, and operationalize, graphic design competencies. This was the purpose of Study 2.

4.2 Literature Specific to this Chapter

The section on skills and domain briefly described the ambitions that lay behind the development of O*NET: to create a method and framework that could identify the prerequisite competencies for every job available in the US labour market. To do so, the development of O*NET engaged a number of seminal academics within the field of vocational psychology, and the work of Fleishman and colleagues is of particular note, specifically their development of the F-JAS (Fleishman Job Analysis Survey).

Similarly, the section on defining skills briefly introduced relevant concepts and language from the field of vocational psychology, and within this job analysis research. The purpose of which was to emphasise the hierarchical role of abilities: that abilities underpin skills, and without requisite abilities the opportunity to develop skills for particular tasks are limited.

The points above are related. O*NET's findings on abilities are directly based on the Fleishman Job Analysis Survey (F-JAS). Within job analysis research F-JAS is a well-established instrument, and consider to be one that is thoroughly researched, with many studies to evidence construct and predictive validity (Wilson, 2007). In its current form, the F-JAS is comprised of a set of 52 abilities that are deemed to be important to competent performance within a wide range of jobs. There is also the F-JAS2 (Fleishman-Post Analyses Survey) that identifies a further 21 interpersonal and social abilities relevant to job performance (Mazilescu, et a.l, 2010). O*NET is based on the first set of 52 abilities. Figure 8, below, shows O*NET's prioritisation (from 0% to 100%) of these abilities in relation to graphic design.

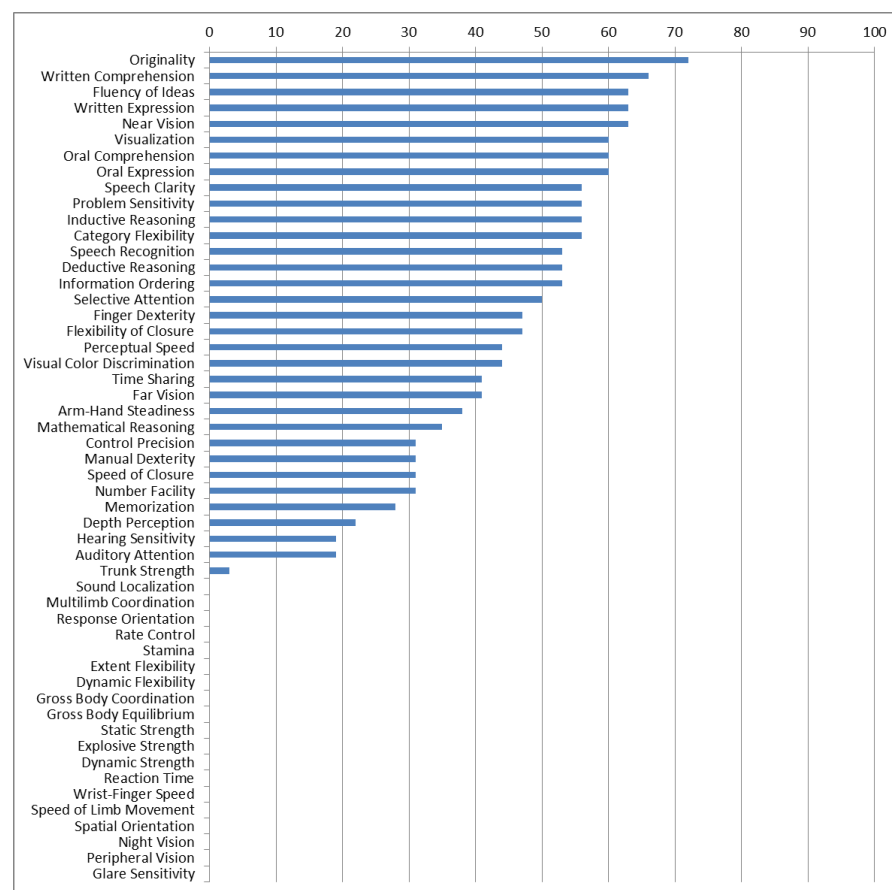


Figure 8: O*NET's prioritisation of graphic design abilities

In turn, the lineage of F-JAS is based on the seminal work of Prof. John B. Carroll's model of cognitive abilities. Carroll's work, considered to be of profound importance in understanding human abilities and intelligence (McGrew, 2009; Jensen, 2004; Horn 1998; Burns 1994), was one of the main models synthesised into Cattell, Horn & Carroll (CHC) theory.

4.2.1 CHC Theory

CHC theory, and the framework it offers for understanding human intelligence, has become a model of choice for many researchers involved in psychometric studies (Keith & Reynolds, 2010), and in recent creativity studies is gaining advocacy within this field of research (Avitia & Kaufman, 2014; Kaufman, 2015; Silvia, 2015).

Based on the synthesis of Cattell–Horn's Gf–Gc (fluid & crystallised intelligence) and Carroll's Three-Stratum model, it proposes that broad intelligence (often labelled as G) appears to be underpinned by at least ten broad abilities and knowledge:

- **Fluid Reasoning (Gf)**. e.g. the ability to solve novel problems
- **Comprehension-Knowledge (Gc)**. e.g. the breadth and depth of cultural knowledge acquired through education and life experiences
- **Visual Processing (Gv)**. e.g. the ability to generate or transform visual images
- **Auditory Processing (Ga)**. e.g. the ability to distinguish and identify sound patterns from background noise
- **Short-Term Memory (Gsm)**. e.g. the ability to maintain awareness of events that have occurred in the last minute.
- **Long-Term Storage and Retrieval (Glr)** e.g. the ability to name a list of items a week after first learning the list.

- **Cognitive Processing Speed** (Gs)
- **Decision and reaction speed** (Gt)
- **Quantitative Knowledge** (Gq)
- **Reading and Writing** (Grw)

A further six broad abilities and knowledge belong to the model, though were tentative at the time of version 1.0 CHC theory (McGrew, 2009). These included:

- Domain-specific knowledge (Gkn),
- Tactile Abilities (Gh),
- Kinesthetic Abilities (Gk)
- Olfactory Abilities (Go)
- Psychomotor abilities (Gp)
- Psychomotor speed (Gps)

Within the CHC model, each broad ability and knowledge above cascades to narrow abilities (sometimes referred to as lower order abilities), of which the model has identified eighty-two such constructs. The hierarchal terminology of these labels could infer that these constructs are lesser, limited, or possibly not as important as broad abilities or G itself, however, this interpretation is argued against (McGrew 2004).

Since its initial synthesis, CHC has continued to develop, and version 2.0 CHC theory had been proposed by Schneider and McGrew (2012) when Study 2 began. The changes between CHC theory version 1.0 and version 2.0 are relatively incremental, which seems beneficial to building on past research as well as future-proofing current findings. For relative simplicity, an edited version of their taxonomy is provided in Table 11: with broad

abilities (on the left) to narrow abilities (in the middle), followed by a brief description of each narrow ability.

Table 11: CHC theory

CHC Broad Ability/Knowledge	CHC Narrow Ability/Knowledge	Description
Fluid Reasoning (Gf)	Induction (I)	The ability to observe a phenomenon and discover the underlying principles or rules that determine its behaviour
Fluid Reasoning (Gf)	General Sequential Reasoning (RG)	The ability to reason logically using known premises and principles
Fluid Reasoning (Gf)	Quantitative Reasoning (RQ)	The ability to reason with numbers, mathematical relations, and operators
Short-Term Memory (Gsm)	Memory Span (MS)	The ability to immediately reproduce information in the same sequence in which it was represented
Short-Term Memory (Gsm)	Attentional Control (WM)	The ability to focus attention while avoiding distracting stimuli
Long-Term Storage & Retrieval (Glr)	Associative Memory (MA)	The ability to remember previously unrelated information as having been paired
Long-Term Storage & Retrieval (Glr)	Meaningful Memory (MM)	The ability to remember narratives and other forms of semantically related information
Long-Term Storage & Retrieval (Glr)	Free Recall Memory (M6)	The ability to recall lists in any order
Long-Term Storage & Retrieval (Glr)	Ideational Fluency (FI)	Ability to rapidly produce a series of ideas, words, or phrases related to a specific condition or object
Long-Term Storage & Retrieval (Glr)	Associational Fluency (FA)	Ability to rapidly produce a series of original or useful ideas related to a particular concept
Long-Term Storage & Retrieval (Glr)	Expressional Fluency (FE)	Ability to rapidly think of different ways of expressing an idea
Long-Term Storage & Retrieval (Glr)	Sensitivity to Problems (SP)	Ability to rapidly think of a number of alternative solutions to a particular practical problem
Long-Term Storage & Retrieval (Glr)	Originality/Creativity (FO)	Ability to rapidly produce original, clever, and insightful responses to a given topic, situation, or task
Long-Term Storage & Retrieval (Glr)	Naming Facility (NA)	Ability to rapidly call objects by their names

Long-Term Storage & Retrieval (Glr)	Word Fluency (FW)	Ability to rapidly produce words that share one or more phonemic or orthographic features
Long-Term Storage & Retrieval (Glr)	Figural Fluency (FF)	Ability to rapidly draw or sketch as many things as possible when presented with a non-meaningful visual stimulus
Long-Term Storage & Retrieval (Glr)	Figural Flexibility (FX)	Ability to rapidly draw different solutions to figural problems
Processing Speed (Gs)	Perceptual Speed (P)	Speed at which visual stimuli can be compared for similarity or difference
Processing Speed (Gs)	Rate-of-Test-Taking (R9)	Speed and fluency with which simple cognitive tests are completed
Processing Speed (Gs)	Number Facility (N)	Speed at which basic arithmetic operations are performed accurately
Processing Speed (Gs)	Reading Speed (fluency) (RS)	Rate of reading text with full comprehension
Processing Speed (Gs)	Writing Speed (fluency) (WS)	Rate at which words or sentences can be generated or copied
Reaction and Decision Speed (Gt)	Simple Reaction Time (R1)	Reaction time to the onset of a single stimulus (visual or auditory)
Reaction and Decision Speed (Gt)	Choice Reaction Time (R2)	Reaction time when a very simple choice must be made
Reaction and Decision Speed (Gt)	Semantic Processing Speed (R4)	Reaction time when a decision requires some very simple encoding and mental manipulation of the stimulus content
Reaction and Decision Speed (Gt)	Mental Comparison Speed (R7)	Reaction time where stimuli must be compared for a particular characteristic or attribute
Reaction and Decision Speed (Gt)	Inspection Time (IT)	The speed at which differences in stimuli can be perceived
Psychomotor Speed (Gps)	Speed of Limb Movement (R3)	The speed of arm and leg movement
Psychomotor Speed (Gps)	Writing Speed (fluency) (WS)	The speed at which written words can be copied
Psychomotor Speed (Gps)	Speed of Articulation (PT)	Ability to rapidly perform successive articulations with the speech musculature
Psychomotor Speed (Gps)	Movement Time (MT)	The time taken to physically move a body part to make a required response
Comprehension-Knowledge (Gc)	General Verbal Information (K0)	Breadth and depth of knowledge that one's culture deems essential,

		practical, or otherwise worthwhile for everyone to know
Comprehension-Knowledge (Gc)	Language Development (LD)	General understanding of spoken language at the level of words, idioms, and sentences
Comprehension-Knowledge (Gc)	Lexical Knowledge (VL)	Knowledge of the definitions of words and the concepts that underlie them
Comprehension-Knowledge (Gc)	Listening Ability (LS)	Ability to understand speech
Comprehension-Knowledge (Gc)	Communication Ability (CM)	Ability to use speech to communicate one's thoughts clearly
Comprehension-Knowledge (Gc)	Grammatical Sensitivity (MY)	Awareness of the formal rules of grammar and morphology of words in speech
Domain-Specific Knowledge (Gkn)	General Science Information (K1)	Range of scientific knowledge (e g , biology, physics, engineering, mechanics, electronics)
Domain-Specific Knowledge (Gkn)	Knowledge of Culture (K2)	Range of knowledge about the humanities (e g , philosophy, religion, history, literature, music, and art)
Domain-Specific Knowledge (Gkn)	Geography Achievement (A5)	Range of geography knowledge (e g , capitals of countries)
Domain-Specific Knowledge (Gkn)	Mechanical Knowledge (MK)	Knowledge about the function, terminology, and operation of ordinary tools, machines, and equipment
Domain-Specific Knowledge (Gkn)	Knowledge of Behavioral Content (BC)	Knowledge or sensitivity to nonverbal human communication/interaction systems (e g , facial expressions and gestures)
Domain-Specific Knowledge (Gkn)	Foreign Language Proficiency (KL)	Similar to Language Development but in another language
Domain-Specific Knowledge (Gkn)	Knowledge of Signing (KF)	Knowledge of finger-spelling and signing (e g , American Sign Language)
Domain-Specific Knowledge (Gkn)	Skill in Lip-Reading (LP)	Competence in the ability to understand communication from others by watching the movement of their mouths and expressions
Reading and Writing (Grw)	Reading Decoding (RD)	Ability to identify words from text
Reading and Writing (Grw)	Reading Comprehension (RC)	Ability to understand written discourse

Reading and Writing (Grw)	Reading Speed (RS)	Rate at which a person can read connected discourse with full comprehension
Reading and Writing (Grw)	Spelling Ability (SG)	Ability to spell words
Reading and Writing (Grw)	English Usage (EU)	Knowledge of the mechanics of writing (e g , capitalization, punctuation, and word usage)
Reading and Writing (Grw)	Writing Ability (WA)	Ability to use text to communicate ideas clearly
Reading and Writing (Grw)	Writing Speed (WS)	Ability to copy or generate text quickly
Quantitative Knowledge (Gq)	Mathematical Knowledge (KM)	Range of general knowledge about mathematics
Quantitative Knowledge (Gq)	Mathematical Achievement (A3)	Measured (tested) mathematics achievement
Visual Processing (Gv)	Visualization (Vz)	The ability to perceive complex patterns and mentally simulate how they might look when transformed
Visual Processing (Gv)	Speeded Rotation (Spatial Relations; SR)	The ability to solve problems quickly using mental rotation of simple images
Visual Processing (Gv)	Closure Speed (CS)	Ability to quickly identify a familiar meaningful visual object from incomplete visual stimuli, without knowing in advance what the object is
Visual Processing (Gv)	Flexibility of Closure (CF)	Ability to identify a visual figure or pattern embedded in a complex distracting visual pattern, when knowing in advance what the pattern is
Visual Processing (Gv)	Visual Memory (MV)	Ability to remember complex images over short periods of time (less than 30 seconds)
Visual Processing (Gv)	Spatial Scanning (SS)	Ability to visualise a path out of a maze or a field with many obstacles
Visual Processing (Gv)	Serial Perceptual Integration (PI)	Ability to recognize an object after only parts of it are shown in rapid succession
Visual Processing (Gv)	Length Estimation (LE)	The ability to visually estimate the length of objects
Visual Processing (Gv)	Perceptual Illusions (IL)	The ability to not be fooled by visual illusions
Visual Processing (Gv)	Perceptual Alternations (PN)	Consistency in the rate of alternating between different visual perceptions

Visual Processing (Gv)	Imagery (IM)	Ability to mentally imagine very vivid images
Auditory Processing (Ga)	Phonetic Coding (PC)	Ability to hear phonemes distinctly
Auditory Processing (Ga)	Speech Sound Discrimination (US)	Ability to detect and discriminate differences in speech sounds under conditions of little or no distraction or distortion
Auditory Processing (Ga)	Resistance to Auditory Stimulus Distortion (UR)	Ability to hear words correctly even under conditions of distortion or loud background
Auditory Processing (Ga)	Memory for Sound Patterns (UM)	Ability to retain (on a short-term basis) auditory events such as tones, tonal patterns, and voices
Auditory Processing (Ga)	Maintaining and Judging Rhythm (U8)	Ability to recognize and maintain a musical beat
Auditory Processing (Ga)	Musical Discrimination and Judgment (U1 U9)	Ability to discriminate and judge tonal patterns in music with respect to melodic, harmonic, and expressive aspects
Auditory Processing (Ga)	Absolute Pitch (UP)	Ability to perfectly identify the pitch of tones
Auditory Processing (Ga)	Sound Localization (UL)	Ability to localize heard sounds in space
Olfactory Abilities (Go)	Olfactory Memory (OM)	Ability to recognize previously encountered distinctive odors
Tactile Abilities (Gh)	?	?
Kinesthetic Abilities (Gk)	?	?
Psychomotor Abilities (Gp)	Static Strength (P3)	The ability to exert muscular force to move (push, lift, pull) a relatively heavy or immobile object
Psychomotor Abilities (Gp)	Multilimb Coordination (P6)	The ability to make quick specific or discrete motor movements of the arms or legs
Psychomotor Abilities (Gp)	Finger Dexterity (P2)	The ability to make precisely coordinated movements of the fingers (with or without the manipulation of objects)
Psychomotor Abilities (Gp)	Manual Dexterity (P1)	Ability to make precisely coordinated movements of a hand or a hand and the attached arm
Psychomotor Abilities (Gp)	Arm-Hand Steadiness (P7)	The ability to precisely and skillfully coordinate arm–hand positioning in space
Psychomotor Abilities (Gp)	Control Precision (P8)	The ability to exert precise control over muscle movements, typically in response to environmental feedback

Psychomotor Abilities (Gp)	Aiming (AI)	The ability to precisely and fluently execute a sequence of eye–hand coordination movements for positioning purposes
Psychomotor Abilities (Gp)	Gross Body Equilibrium (P4)	The ability to maintain the body in an upright position in space or regain balance after balance has been disturbed

Given the fundamental influence CHC theory has had on the theoretical development of human abilities, and the practical development of psychometric testing, it would appear that CHC theory remains at present a comprehensive and coherent theoretical model. As mentioned above CHC theory has an impressive status within abilities research. Whilst it continues to adapt and evolve CHC theory does appear to be the most well-regarded theoretical model of intelligence and human abilities to date; has held this position for some time, and appears to have longevity (as much as any seminal theory can be predicted to have longevity).

The caveats that proponents of CHC theory espouse are not to treat the model as static, but rather to see this as a useful taxonomy of abilities and knowledge upon which further research in the field of intelligence and human abilities can build. McGrew (2009) acknowledges that his 1997 work to synthesise Cattell & Horn with Carroll was partly based on pragmatic reasons as much as empirical findings and thus, needs to be seen in this light: it was an attempt to classify and give a degree of order to the variety of constructs and terminology within intelligence research at the time.

As discussed previously, arguably, parallels can be made between the rationale to synthesise intelligence research that resulted in CHC theory, and the complexity of information regarding graphic design competencies at the time of this study.

For this reason, the suggestion is that CHC theory may also offer a suitable conceptual framework within which to classify existing models of graphic design competencies. Or, put more succinctly: Is CHC theory a useful coding framework to map graphic design competencies? To date, it appears no other researcher or organisation involved in defining graphic design competencies has undertaken this approach to classification.

The value of such *mapping* is in its utility to clarify overlaps, commonalities, and genuine areas of distinction across models; it may be possible that CHC theory could underpin the theoretical silos that appear to have occurred within the graphic design competency models explored in this study.

Focus on CHC theory could also offer a number of other potential benefits. Whilst CHC theory is not a static framework nor dogmatic in its ethos; theoretically, it is supposed to be comprehensive. As such, graphic design competencies that do not readily map to CHC Theory are important: either they require CHC theory to adapt, and acknowledge their inclusion, or suggest a more objective basis on which to question their relative importance.

This last point is worth considering further, as a more analytical approach is likely to be valuable. It is useful to emphasise that decisions on what competencies to include or disregard, for many industries, is often premised on the views of a working group. Yet, previous studies have shown that the influence of dominant individuals upon a group dynamic can significantly impact on the contribution of other group members (Robson, 1993). Practically, that may mean that some competencies are given greater importance on the basis of the individual preference of a working group member, whilst others are overlooked.

This theme of a dominant view can also be seen in the composition of working groups. For example, one of the criticisms Macdonald (2006) had of The Film Skills Group (2003), was that it had no representation from educators or researchers on the steering group (the steering group was made of 25 individuals from a variety of domains, but within this group over a quarter of individuals worked directly for either the Film Council or Skillset).

It is clear from the research methodology section of the final report (Film Skills Group, 2003) that representation from the educational sector was considered later in the study: of the 338 people consulted, 60 were classified as belonging to education and training. However, the selection of education providers "on the basis of recommendations from the project steering group, stakeholders and practitioners" (p.21) lends weight to Macdonald's concern of bias. Aside from this issue, gathering 338 participants to take part in such a study appears impressive, and a clear attempt to overcome individual bias on skills. On closer inspection, the number of participants varied depending on the sector. For example, eleven film industry sectors were classified for the study, within these only two individuals took part from the commercials sector, and four participants took part from the video sector. Given such low numbers, two participants for a sector is problematic and likely to skew the findings on what the skills for those sectors may be.

The point to highlight is that a conceptual framework like CHC theory may offer a means to stand back from the individual interests of informants, research participants and working group members, and explore their perceptions within an overall model of human abilities.

Aside from this analytical value, where competencies do map to CHC Theory, it will be possible to track back to the psychometric tests associated with a narrow ability; in this way identifying the operational means to test specific abilities. Moreover, being able to link language used within the domain of graphic design within the terminology of vocational psychologist should allow access to relevant research studies and literature not previously associated with each other.

4.3 Method

The approach to mapping was built upon methods of analysis associated with textual coding undertaken in quantitative studies, particularly grounded theory. In this respect, coding and mapping was a process of identifying descriptions given for graphic design competencies and matching these, if possible, with the descriptions of narrow abilities in CHC Theory.

To manage the wealth and detail of information available, even with the focus on four specific models of graphic design competencies, the range of abilities under analysis was refined. O*NET's job analysis of graphic designers had identified 19 of the 52 abilities in its model as not relevant to graphic design (Night vision, or Sound Localization for example). In terms of refining the range of abilities, this degree of prioritisation offered an informed basis upon which abilities to exclude from the study.

Additionally, Study 2 would focus on cognitive abilities, rather than physical, psychomotor or sensory abilities. Within the O*NET model, for example, cognitive abilities account for 69% of the ratings given to graphic designers, in contrast, physical abilities only account for 0.2%; sensory abilities just over 21%, and psycho-motor abilities almost 10%. Two sensory abilities are worth a specific mention at this point. Near Vision, for example,

ranked within the top 20% of important abilities; speech clarity also ranked highly. As cognitive abilities were the area of interest for this study, while O*NET considers these sensory abilities important for job performance within graphic design they were discounted from the initial mapping research.

For the final analysis, this left a prioritised list of ten abilities suggested as highly important to graphic design competency, and each classifies as cognitive in nature. In order of importance, they were:

- Originality
- Written Comprehension
- Fluency of Ideas
- Written Expression
- Visualisation
- Oral Comprehension
- Oral Expression
- Problem Sensitivity
- Inductive Reasoning
- Category Flexibility

As a further consideration of method, some areas of CHC Theory gave a more detailed description and finer distinctions on abilities than O*NET. CHC language and terminology (while relatively accessible) had been worded for clinicians and those involved in psychometric study (Schneider & McGrew, 2012). In contrast, O*NET's ability descriptions had been worded for job analysis professionals, and to be assessable to incumbents (research participants, if you will, in specific occupations that took part in O*NET's job analysis: in this case graphic designers). For this reason, it was prudent to check for comparable descriptions between CHC and O*NET on the ten abilities described above, before mapping these abilities to each graphic design competencies model in this study, namely: O*NET's Graphic designer core tasks; NASAD's Graphic Design (Essential Competencies); BTEC Specialist: Graphic Design, and BTEC Higher National in Graphic Design.

With regard to BTEC's Specialist qualification, a final consideration of method was that some units were more specific than other and relevant to specialist domains within, or related to, graphic design. As described in section 2.6.3, on the relationship between skills and domain, BTEC's specialist qualification in graphic design was intended to provide the competencies required for preparation to work within graphic design. Course providers could choose four units from a number of options, and in this way tailor their course to local needs or markets. Given this range of options, for this study, not all of the units were selected for mapping. Table 12 shows those units not included and in brackets the associated specialism/domain

Table 12: BTEC Specialist units not included in study 2

<ul style="list-style-type: none"> • Information Graphics (Signage) • Website Design (web design) • Design for Advertising (Advertising) • Graphics for 3D Application (3D modelling/Games Art) • Digital Story Telling (animation) • Specialist Illustration Computer Applications (illustration) • Factual Writing in Art and Design (Copywriting)

Accounting for the exclusions above, the following units were included, and the learning outcomes for each unit are described in Table 13 (BTEC, 2010).

Table 13: BTEC Specialist units, and learning outcomes, included in study 2

Name of Unit	Learning Outcome
Typographic and Layout Design	Understand typographic and layout conventions used in the work of others
	Know terminology used within typographic and layout design
	Be able to use computers and other media to develop typographic and layout designs
	Be able to evaluate own typographic and layout design outcomes.
Words and Images in Graphic Design	Know how words and images are used in the work of others
	Understand the relationship of words and images when developing ideas
	Be able to produce final outcomes that explore the potential of words and images.
	Know about graphic image making
	Be able to develop ideas for graphic image making

Graphic Image Making	Be able to use media safely to produce graphic image outcomes
	Be able to present own design outcomes against a given theme or assignment brief.
Narrative Image Making	Be able to investigate examples of narrative imagery
	Understand the processes used in illustrating narrative
	Be able to develop ideas for narrative imagery in response to given themes
	Be able to present narrative imagery.
Typefaces and Letter Forms	Understand historical and contemporary developments of typefaces and letter forms
	Know how typeface styles and letter forms communicate a message
	Know correct terminology for typefaces and letter forms
	Be able to create design work that explores typefaces and letter forms.
Mixed Media Image Making	Be able to use experimental and non-traditional graphic media techniques and processes
	Be able to use mixed media in producing graphic image outcomes
	Be able to select reproduction techniques towards graphic image outcomes.
Graphic Media, Techniques and Technology	Know about graphic media, techniques and technology
	Be able to use graphic media, techniques and technology within own design work
	Be able to present design outcomes using graphic media, techniques and technology.

4.4 Results

4.4.1 Mapping ONET Graphic Design Abilities to CHC Theory

For the most part, each O*NET ability mapped across to CHC theory narrow abilities (see Table 14). Problem Sensitivity, and Category Flexibility, however, did not.

Firstly, problem sensitivity was described in O*NET as "The ability to tell when something is wrong or is likely to go wrong. It does not involve solving the problem, only recognising there is a problem." In contrast, (though similarly titled) Sensitivity to Problems/Alternative Solution Fluency (SP) was described in CHC theory as the "Ability to rapidly think of a number of alternative solutions to a particular practical problem". This does appear confused as little of the description is related to problem sensitivity, and the majority of the description is focused on developing alternative solutions to a problem. In this respect, the

description has much more in common with Ideational Fluency (FI) in CHC, and Fluency of Ideas in O*NET. Yet, sensitivity to problems, particularly within creativity strategies such as Creative Problem Solving is considered a key phase. Given this, O*NET's description of Problem Sensitivity appears the most accurate and was the description used in this study.

Secondly, Category Flexibility, described in the O*NET model as "The ability to generate or use different sets of rules for combining or grouping things in different ways", did not easily map across to a distinct CHC ability other than Induction (I). Yet, the O*NET model had identified Inductive Reasoning as an important graphic design ability and, as described (see Table 14), both CHC and O*NET's titles and descriptions appear synonymous. Given this, Category Flexibility was removed from the coding framework, and subsumed under Induction (I) and Inductive Reasoning for the analysis.

*Table 14: Mapping of CHC to O*NET*

CHC Narrow Ability	Description	O*NET	Description
Induction (I)	The ability to observe a phenomenon and discover the underlying principles or rules that determine its behaviour	Inductive Reasoning	The ability to combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events).
Ideational Fluency (FI)	Ability to rapidly produce a series of ideas, words, or phrases related to a specific condition or object	Fluency of Ideas	The ability to come up with a number of ideas about a topic (the number of ideas is important, not their quality, correctness, or creativity).
Originality/ Creativity (FO)	Ability to rapidly produce original, clever, and insightful responses to a given topic, situation, or task	Originality	The ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem.
Listening Ability (LS)	Ability to understand speech	Oral Comprehension	The ability to listen to and understand information and ideas presented through spoken words and sentences.

Communication Ability (CM)	Ability to use speech to communicate one's thoughts clearly	Oral Expression	The ability to communicate information and ideas in speaking so others will understand.
Reading Comprehension (RC)	Ability to understand written discourse	Written Comprehension	The ability to read and understand information and ideas presented in writing.
Writing Ability (WA)	Ability to use text to communicate ideas clearly	Written Expression	The ability to communicate information and ideas in writing so others will understand.
Visualization (Vz)	The ability to perceive complex patterns and mentally simulate how they might look when transformed	Visualization	The ability to imagine how something will look after it is moved around or when its parts are moved or rearranged.

4.4.2 ONET/CHC Cross References with Graphic Design Competencies

Based on these nine abilities and relevant descriptions, each ability was cross-referenced against each specific competency identified through the four graphic design models selected for this study: BTEC Specialist, BTEC Higher Nationals, NASAD, and O*NET graphic design tasks. Following the illustration given by Peterson et al. (2001, p.479) on the development of O*NET rating scales, a guiding principle was to focus on the importance of an ability to a competency. Each combination was considered through the question: How important is ability "A" to performance on competency "B". Where importance was considered to be high, the number 1 was coded to denote this, where it was not considered to be of high importance for that particular competency the field was left blank. The tables 15 to 19, below, show the results for each model, with frequency counts provided at the bottom of each matrix.

As each model had a different number of competencies (for example NASAD had 6, and BTEC Specialist had 25) using frequency scores to compare would be skewed in favour of those models with more competencies. To mitigate this, frequency scores were ranked to show the relative overall importance for each of the nine abilities across each model

(1 was the least rank, and 9 the highest).

Table 15: O*NET Core Tasks cross-referenced to abilities

	Originality	Written Comprehension	Written Expression	Fluency of Ideas	Oral Comprehension	Oral Expression	Visualization	Problem Sensitivity	Inductive Reasoning
O*NET Core Tasks									
Create designs, concepts, and sample layouts based on knowledge of layout principles and aesthetic design concepts.	1	1		1	1	1	1	1	1
Determine size and arrangement of illustrative material and copy, and select style and size of type.	1	1	1	1				1	1
Confer with clients to discuss and determine layout design.	1	1			1	1		1	1
Develop graphics and layouts for product illustrations, company logos, and Internet websites.	1	1		1	1	1		1	1
Review final layouts and suggest improvements as needed.	1	1	1	1	1	1		1	1
Prepare illustrations or rough sketches of material, discussing them with clients or supervisors and making necessary changes.	1	1		1	1	1		1	1
Use computer software to generate new images.	1	1						1	1
Key information into computer equipment to create layouts for client or supervisor.		1			1			1	
Maintain archive of images, photos, or previous work products.		1	1						
Prepare notes and instructions for workers who assemble and prepare final layouts for printing.		1	1			1		1	1
Draw and print charts, graphs, illustrations, and other artwork, using computer.	1	1						1	
Study illustrations and photographs to plan presentation of materials, products, or services.		1						1	
Research new software or design concepts.		1							1
Total	8	13	4	5	6	6	9	10	9

ONET	Total	Rank
Written Expression	4	1
Fluency of Ideas	5	2
Oral Comprehension	6	3.5
Oral Expression	6	3.5
Originality	8	5
Visualization	9	6.5
Inductive Reasoning	9	6.5
Problem Sensitivity	10	8
Written Comprehension	13	9

Table 16: NASAD essential competencies cross referenced to abilities

	Originality	Written Comprehension	Written Expression	Fluency of Ideas	Oral Comprehension	Oral Expression	Visualization	Problem Sensitivity	Inductive Reasoning
NASAD Graphic Design (essential competencies)									
The ability to solve communication problems...	1	1	1	1	1	1	1	1	1
The ability to describe and respond to the audiences and contexts which communication solutions must address...	1	1	1		1	1		1	1
The ability to create and develop visual form in response to communication problems...	1	1		1	1	1	1	1	1
An understanding of tools and technology...		1			1			1	
An understanding of design history, theory, and criticism from a variety of perspectives...		1	1	1	1	1		1	1
An understanding of basic business practices...					1	1		1	
Total	3	5	3	3	6	5	2	6	4

NASAD/CHC	Total	Rank
-----------	-------	------

Visualization	2	1
Originality	3	3
Written Expression	3	3
Fluency of Ideas	3	3
Inductive Reasoning	4	5
Written Comprehension	5	6.5
Oral Expression	5	6.5
Oral Comprehension	6	8.5
Problem Sensitivity	6	8.5

Table 17: BTEC Specialist cross-referenced to abilities

Unit		Originality	Written Comprehension	Written Expression	Fluency of Ideas	Oral Comprehension	Oral Expression	Visualization	Problem Sensitivity	Inductive Reasoning
Unit	BTEC Specialist: Graphic Design (learning outcomes)									
Typographic	1 Understand typographic and layout conventions used in the work of others	1					1	1	1	
Typographic a	2 Know terminology used within typographic and layout design	1					1			
Typographic a	3 Be able to use computers and other media to develop typographic and layout designs							1		
Typographic a	4 Be able to evaluate own typographic and layout design outcomes.			1			1	1	1	
Words and Im	1 Know how words and images are used in the work of others	1					1	1		1
Words and Im	2 Understand the relationship of words and images when developing ideas	1					1	1		1
Words and Im	3 Be able to produce final outcomes that explore the potential of words and images.	1	1	1			1			
Graphic Image	1 Know about graphic image making	1					1			
Graphic Image	2 Be able to develop ideas for graphic image making	1		1			1			
Graphic Image	3 Be able to use media safely to produce graphic image outcomes									
Graphic Image	4 Be able to present own design outcomes against a given theme or assignment brief.	1	1	1			1	1		
Narrative Imag	1 Be able to investigate examples of narrative imagery							1		
Narrative Imag	2 Understand the processes used in illustrating narrative			1			1	1		
Narrative Imag	3 Be able to develop ideas for narrative imagery in response to given themes	1	1	1			1	1		
Narrative Imag	4 Be able to present narrative imagery.	1					1			
Typefaces and	1 Understand historical and contemporary developments of typefaces and letter forms	1					1		1	
Typefaces and	2 Know how typeface styles and letter forms communicate a message	1					1	1		
Typefaces and	3 Know correct terminology for typefaces and letter forms	1					1			
Typefaces and	4 Be able to create design work that explores typefaces and letter forms.	1	1	1			1	1		
Mixed Media I	1 Be able to use experimental and non-traditional graphic media techniques and processes							1	1	
Mixed Media I	2 Be able to use mixed media in producing graphic image outcomes							1		1
Mixed Media I	3 Be able to select reproduction techniques towards graphic image outcomes.									1
Graphic Media	1 Know about graphic media, techniques and technology			1			1			
Graphic Media	2 Be able to use graphic media, techniques and technology within own design work							1		
Graphic Media	3 Be able to present design outcomes using graphic media, techniques and technology.			1			1			
		6	8	8	5	0	15	16	5	4

BTEC Specialist Graphic Design	Total	Rank
Oral Comprehension	0	1
Inductive Reasoning	4	2
Fluency of Ideas	5	3.5
Problem Sensitivity	5	3.5
Originality	6	5
Written Comprehension	8	6.5
Written Expression	8	6.5
Oral Expression	15	8

Visualization	16	9
---------------	----	---

Table 18: BTEC HN's cross-referenced to abilities

		Originality	Written Comprehension	Written Expression	Fluency of Ideas	Oral Comprehension	Oral Expression	Visualization	Problem Sensitivity	Inductive Reasoning
Unit	BTEC Higher National in Graphic Design (learning outcomes)									
Visual Communic	1 Be able to communicate ideas and concepts by researching visual techniques			1	1		1	1		1
Visual Communic	2 Be able to select visual communication techniques to realise creative intentions	1			1				1	1
Visual Communic	3 Be able to produce work which demonstrates the use of visual communication			1	1				1	1
Visual Communic	4 Understand the potential for personal development through the application of new approaches to visual communication.			1			1			
Ideas Generation	1 Be able to exploit idea-generating techniques	1			1			1		1
Ideas Generation	2 Be able to implement complex ideas visually	1			1				1	1
Ideas Generation	3 Be able to express ideas using visual techniques	1			1				1	1
Ideas Generation	4 Understand the cultural contexts in which ideas are visually transmitted.		1	1			1			1
Contextual and C1	1 Be able to undertake in-depth research		1	1	1		1		1	1
Contextual and C2	2 Understand influences on art and design activities and outcomes through the interpretation and analysis of		1	1			1		1	1
Contextual and C3	3 Be able to assess, interpret and evaluate information		1	1			1		1	1
Contextual and C4	4 Be able to evaluate and present conclusions.			1			1		1	1
Professional Prac	1 Be able to place themselves and their work in the context of their selected discipline			1			1			1
Professional Prac	2 Understand their specialist area and the career opportunities available			1			1			1
Professional Prac	3 Be able to develop and present a professional portfolio in an appropriate format	1		1	1		1	1	1	1
Professional Prac	4 Understand how to promote themselves and their work professionally.			1			1			1
Project Design, Ir	1 Be able to formulate a project			1						1
Project Design, Ir	2 Be able to implement the project within agreed procedures and to specification		1							1
Project Design, Ir	3 Be able to evaluate the project outcomes			1			1			1
Project Design, Ir	4 Be able to present the project outcomes.			1			1			
		5	5	15	8	0	13	7	9	15

BTEC HN's	Total	Rank
Oral Comprehension	0	1
Originality	5	2.5
Written Comprehension	5	2.5
Visualization	7	4
Fluency of Ideas	8	5
Problem Sensitivity	9	6
Oral Expression	13	7
Written Expression	15	8.5
Inductive Reasoning	15	8.5

4.4.3 Overall Ranking and Priorities.

Ranks were combined across each of the four frameworks to give an overall indicator of relative importance for each of the nine abilities. Table 19 shows the total rank score (one was the lowest rank, and nine the highest).

Table 19: Comparison of abilities across each model

	BTEC Specialist	O*NET	Higher Nationals	NASAD		Total Score	Total Rank
Fluency of Ideas	3.5	2	5	3		13.5	1
Oral Comprehension	1	3.5	1	8.5		14	2
Originality	5	5	2.5	3		15.5	3
Written Expression	6.5	1	8.5	3		19	4
Visualization	9	6.5	4	1		20.5	5
Inductive Reasoning	2	6.5	8.5	5		22	6
Written Comprehension	6.5	9	2.5	6.5		24.5	7
Oral Expression	8	3.5	7	6.5		25	8
Problem Sensitivity	3.5	8	6	8.5		26	9

4.5 Discussion

The research challenge at the outset of this study was to navigate the level of detail available in O*NET's findings for graphic design abilities with that of organisations more directly associated with graphic design training and education (BTEC and NASAD). By undertaking the method described above, the nine abilities identified by O*NET's graphic design job analysis did appear to map to one or more of the graphic design tasks, skills, and competencies described across the range of models explored. In order of importance, these were:

- 1st Problem Sensitivity
- 2nd Oral Expression
- 3rd Written Comprehension
- 4th Inductive Reasoning
- 5th Visualisation
- 6th Written Expression
- 7th Originality
- 8th Oral Comprehension
- 9th Fluency of Ideas

As a method for condensing the detail and complexity related to differing models of graphic design competencies, CHC theory seems to offer a useful coding framework to map graphic design competencies. Moreover, each of these nine narrow abilities

appeared to cluster around three broad themes related to creativity, language, and visualisation.

For example, Problem Sensitivity, Inductive Reasoning, Originality and Fluency of Ideas, both in title and description, could fit with particular models of creativity. Creative Problems Solving's (CPS) Problem-finding and Idea-finding stages (Isaksen & Treffinger, 2004), or Puccio's FourSight model related to Clarify and Ideate stages suggest themselves (Puccio, 2002). Equally, a further cluster of abilities can be inferred through Oral Expression, Written Comprehension, Written Expression, and Oral Comprehension; suggestive of language (both written and oral) as important to graphic design. Lastly, visualisation, which seems to stand in isolation; or could align with perceptions of creativity: for graphic design researchers like Bennet, visualisation is "a crucial part of creative problem solving in the design process" (2015, p321).

Showing both similarity and difference, returning to *Multiple Intelligences and Graphic Design Ability in Five North Carolina Community Colleges*, Mackie's 2005 study found linguistic intelligence to be more important to graphic design ability than visual-spatial intelligence. With the implication that greater focus on graphic design education should be placed on the development of reading and writing abilities, and less emphasis on drawing ability. Mackie himself was surprised by this aspect of his research, stating "the ability to recognize and understand shape, form, and space is essential to the visual arts, particularly graphic design" (p.85) but, given the mapping process above, perhaps the fundamental importance of written expression and written comprehension for graphic design competence need not be too surprising. Indeed, the combination of image and text can be considered a defining feature of graphic design.

For these reasons, written comprehension and visualisation appeared important constructs to explore further. Equally, practicalities and ethics of research design meant that administering a wide battery of tests for all nine abilities identified was not possible in one study, or desirable. It was considered that two constructs could be run with participants before fatigue and increased attrition rate were likely to impact on findings. The maximum timescale been, for the majority of participants, to be able to undertake all psychometric testing within an hour, or less. As a result, identifying suitable measures and consideration of time requirements for written comprehension and visualisation tests were required.

4.5.1 Operational Measures for Written Comprehension and Visualisation

While CHC broad abilities like Visual Processing (Gv) and Reading and Writing (Grw) may underpin graphic design creativity, the challenge remained how do researchers test these relationships? Mackie's work is closest to an empirical study on the relationship between graphic design creativity and other domain competencies, however, as will be detailed in chapter 5 and 6, the assessment of graphic design creativity was more complex and constrained than Mackie's (2005) research had identified.

Not discounting this need for an operational definition of graphic design creativity, the question of how to operationalise other graphic design abilities was also a focus of this study. In this respect, the mapping of CHC theory to graphic design competency models appeared useful not only to prioritise abilities but to operationalise them for experimental research.

As described previously in this chapter 2 (section 2.6.1), the development of O*NET engaged a number of seminal academics within the field of vocational psychology, and

the work of Fleishman and colleagues was of particular note, specifically their development of the F-JAS (Fleishman Job Analysis Survey). O*NET is based on the first set of 52 abilities identified in F-JAS. In turn, the lineage of F-JAS is based on the seminal work of Prof. John B. Carroll's model of cognitive abilities, later to be integrated into CHC theory. Thus, there is understandable overlap and areas of nuance between O*NET, F-JAS and CHC theory. Each model offered insights for this thesis, but with the focus on testing the specific cognitive abilities identified, it was the details of the F-JAS model that became of greatest use, specifically, Fleishman and Reilly's *Handbook of Human Abilities* (1992). Since first publication, this seminal work has remained the recommended text for researchers applying the F-JAS (Caughron & Mumford, 2012). Within it, each of the 52 abilities that form the F-JAS model are defined and followed by test suggestions with which to measure/operationalize each ability. The practical relevance of this information to operationalising written comprehension and visualisation are discussed below.

4.5.2 Operational Measures for Written Comprehension

Fleishman & Reilly's chapter on reading/written comprehension defines written comprehension as "...the ability to understand written sentences and paragraphs. This ability involves reading and understanding the meaning of words, phrases, sentences and paragraphs. It involves reading; it does not involve writing, listening to, or understanding spoken information" (Fleishman & Reilly, 1992, p.8). The chapter highlights five particular tests:

- Guilford-Zimmerman Aptitude Survey: Verbal Comprehension
- Advanced Vocabulary Test ii –V5
- Understanding Communication
- The PSI basic skills tests for business, industry and government:
Reading Comprehension –BST #2

- Nelson-Denny Reading Test: Forms E and F

To evaluate the current relevance of each test, literature searches were undertaken to consider the use of these test since 2010 within academic journals; two test came to the fore: The Nelson-Denny Reading Test, and Advanced Vocabulary Test ii –V5

The Nelson-Denny Reading Test (NDRT) had at least four journal papers since 2014 and was cited in 85 journal papers from 2013. The test was also used as part of the admission criteria for some American colleges (Molloy College, 2014).

With the "Advanced Vocabulary Test", searches found 8 journal papers in 2014, and from 2010, 43 citations. Available through the Educational Testing Service (ETS), as a part of the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976), the test is comprised of 36 items, each a single word; participants are provided with five other words, and asked to select the one with the closest meaning to the test item (Shaw et al., 2010). The test has two options, V4 and V5 (V4 appeared difficult, but not as difficult as V5).

Alongside current usage, a further consideration for operational use was the time required to administer each test. The Advanced Vocabulary Test was suggested to take 8 minutes (in practice, with reading and understanding the instructions and accounting for the gap between part 1 and part 2, this was more likely to be in the region of 10 to 11 minutes).

The standard time for the Nelson-Denny Reading Test was 35 minutes. However, previous researchers have used only the written comprehension test in their studies (Peverly, Garner, & Vekaria, 2014); this component of the Nelson-Denny has a standard administration time of 20 minutes; it can also be administered through an extended-time

option, and this was suggested to be suitable for populations where English was not their first language, or for institutions where adult learners form the majority of their cohorts. A third administration option is where researchers have shortened the standard time to increase the variability of test scores: the recommended time, in this case, is 15 minutes (Peverly & Sumowski, 2012; Peverly et al., 2013; Peverly, Garner, & Vekaria, 2014). Thus, the administration can span from 15 to 32 minutes.

For the purpose of this thesis, the standard 20 minutes for the written comprehension component seemed most appropriate. While none of the participants would belong to groupings suggested for the extended-time administration, the increased pressure of using the shortened time of 15 minutes was likely to be detrimental given the high prevalence of dyslexia within design education and the design profession more broadly, and posed an additional challenge for a domain unfamiliar with timed examinations and psychometric assessment. Standard timing would also allow the analysis of results relative to the norm tables available.

For the most part, the longstanding use of the Nelson-Denny has occurred through studies based on American populations. A key concern was, given differences between American and British written conventions, would this written comprehension test be suitable for a UK population? The research work of Prof. Jackie Masterson (Professor of Psychology of Reading at UCL's Institute of Education) suggested it would. Her 2004 study (Masterson & Hayes) was based upon a sample of 197 UK students studying either science, social science or humanities degrees, and within a range of year groups. The findings highlighted that UK sample scores were higher than the norms suggested in the Nelson-Denny standardised data, but the authors highlighted that the reasons for this could be attributed to different admissions process between USA and UK universities.

They concluded that "the Nelson-Denny test is appropriate for use with UK university students" (p.34). In subsequent studies, Masterson and Hayes (2007) continued to use the Nelson-Denny as a measure of written comprehension with UK adults.

4.5.2.1 Alternative Measures

Measures not mentioned in Fleishman and Reilly (1992) handbook, but highlighted in research related to written comprehension are the Test of English as a Foreign Language (TEOFL), and the International English Language Testing System (IELTS) (Rashidi & Khosravi, 2010; McKee, 2012). For example, the IELTS academic reading test is a timed test that requires 60 minutes to complete. Three sections of reading material are provided, and participants are given multiple choice answers to select from. One of the concerns from previous research studies, however, has shown that test takers with high levels of written comprehension can gain lower scores on multiple choice formats: partly this appears to be due to seeing subtle ambiguities in set question, and levels of complexity in the answers: in effect, over analysis. Other researchers have suggested that despite these concerns when multiple choices tests are well developed they can offer an appropriate measure (Epstein et al., 2002; McKee, 2012).

Somewhat confusingly a number of tests exist that appear, given their titles, to measure written comprehension but are in fact not used for that purpose. These tests are the Wechsler Test of Adult Reading (WTAR), The National Adult Reading Test (NART) and the Hopkins Adult Reading Test (HART). Each of these can be administered in a relatively short period of time (for example the WTAR takes 10 minutes), and they are each premised upon how a participant pronounces an irregular spelt words: for example words like "heir", "superfluous" or "synecdoche". In some tests, like the WTAR, participants' responses are compared against audio recordings of acceptable variations of pronunciation provided by Wechsler. The major use of these test is within a clinical

context (cognitive damage, for example) and is more often used to predict full-scale IQ scores prior to the onset of illness (the WTAR can be normed against Wechsler's Adult Intelligence Scale), than as a measure of written comprehension.

4.5.2.2 Suitable Measure for Written Comprehension

Given this range of options, the Nelson-Denny Reading Test (form G) appeared to offer a suitable measure of written comprehension. It was a test relevant for adults and college students, took 20 minutes to complete, had a long-standing history and has remained in current use as a measure of written comprehension. Other authors have highlighted the relevance of vocabulary to written comprehension (McKee, 2012), and indeed the full Nelson-Denny Reading Test does include a vocabulary component. However, compared to the Advanced Vocabulary Test, the Nelson-Denny takes a further 7 minutes to complete (Masterson & Hayes, 2004). Given this, the Advanced Vocabulary Test also appeared to be a suitable secondary measure of written comprehension.

4.5.3 Operational Measures for Visualisation

The chapter on visualisation in Fleishman and Reilly's "Handbook of Human Abilities" (1992) defines visualisation as:

the ability to imagine how something will look when it is moved around or when its parts are moved or rearranged. This ability requires the forming of mental images of how patterns or objects would look after certain changes, such as unfolding or rotation. One has to predict how an object, set of objects, or pattern will appear after the changes have been made (p.33).

The chapter then highlights three particular tests:

- Form Board Test (VZ-1)
- Paper Folding (VZ-2)

- Technical Test Battery: Spatial Recognition (TTB:ST9)

To evaluate the current relevance of each test, literature searches were undertaken to consider the use of these test since 2010 within academic studies.

For the Form Board Test (VZ-1) three sources were found, each of which used the test as an operational measure of visualisation. It must be noted that the Form Board Test (VZ-1) is a specific test by the Educational Testing Services, but that the concept of a form board test to measure visualisation does occur in several other formats, such as the Revised Minnesota Paper Form Board Test, which also has a long history and continued usage. This format was also used by Mackie (2005) as his operational measure of visualisation for graphic design students.

The basis of these tests is that users are given a target shape that is to be achieved through the combination of smaller shapes, all these shapes are shown on a sheet of paper. Smaller shapes may be rotated to achieve the target shape, but this is something that the test taker is required to do through mental visualisation: additionally, the visualisations required are two-dimensional in form. The test also requires that these shapes are repositioned and combined through mental visualisation. Two practice examples are given below (Figure 9). On the left is the Form Board Test that requires test takers to decide which combination of the five smaller shapes can be used to make the larger shape. On the right is the Revised Minnesota Paper Form Board Test where test takers are given one set of smaller shapes and asked to find which of the five larger shapes could be achieved through combination (in this example the answer is A). The tests items become progressively more complex and challenging, building from these basic examples shown below.

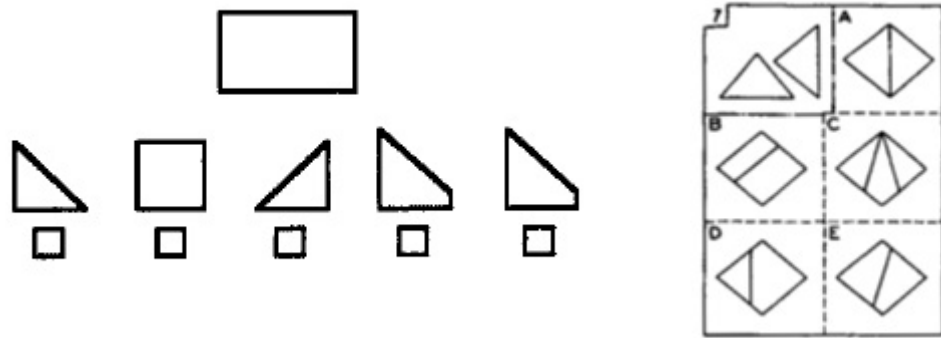


Figure 9: (Left) Form Board Test, (Right) Revised Minnesota Paper Form Board Test

The Paper Folding (VZ-2) is also part of the Educational Testing Services' (ETS) Kit of Factor Referenced Cognitive Tests (Ekstrom et al. 1976) alongside the Form Boards Test (VZ-1). The literature search identified three sources since 2010 that had used this measure in their studies. The test required participants to visualise a piece of paper folded into a particular shape, which then has a hole punched through the folded paper. The task is to determine what the paper will look like unfolded, and where the holes will be positioned. A practice example is provided below (Figure 10): on the left-hand side is the paper, the direction of the fold, and where the hole is made. On the right-hand side are five options suggesting what the paper may look like after unfolding (the answer, in this example, is C).

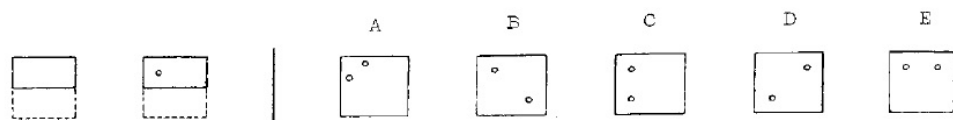


Figure 10: The Paper Folding Test (VZ-2)

For the third test, the Technical Test Battery (TTB): Spatial Recognition, no source were identified post 2010. For the Technical Test Battery, five papers could be found, but these did not appear to use the spatial recognition test as part of their studies. However, in the

UK, a current version of the test was available via Psytec UK (a subsidiary of Psytec International), and the ability to visualise in 2D and 3D belonged under the Spatial Reasoning test; one of the three other measures (Mechanical Reasoning, and Visual Acuity) that make up the TTB2.

4.5.3.1 Alternative Measures

Measures not included in Fleishman and Reilly's chapter on visualisation can be found in the 2013 proceedings of the Engineering Design Graphics Division (EDGD) conference. Kelly et al. (2013) presented a study on visualisation tests for an introductory graphic communication course, of which part of their findings identified 24 tests used by the membership of the EDGD. The top three tests were: the Mental Cutting Test (MCT) developed in 1939 by the College Entrance Examination Board; Vandenberg & Kuse's (1978) Mental Rotations Test (MRT), and the Purdue Spatial Visualization Test: Visualization of Rotations (PSVT:VR) (Guay, 1977).

With regard to the Purdue Spatial Visualizations Test, when first developed the test had three component, one of which was directly related to mental rotations. This component was later extended to become the PSVT:R and as such has longstanding use as a measure of spatial abilities for engineering students undertaking their first year of graphics training (Sorby, 2009). Additionally, the PSVT:R has been described as a "gold standard" for assessment purposes in STEM (Science, Technology, Engineering and Maths) subjects (Sorby et al., 2013; Waller & Lourenco, 2010). The tests consist of 30 items. Within each item is an object that is rotated, and another object that the test takers is required to visualise rotated in the same way as the object above. Test takers are required to choose from five options on how the target rotation would look; an example of which is given below (Fig. 11).

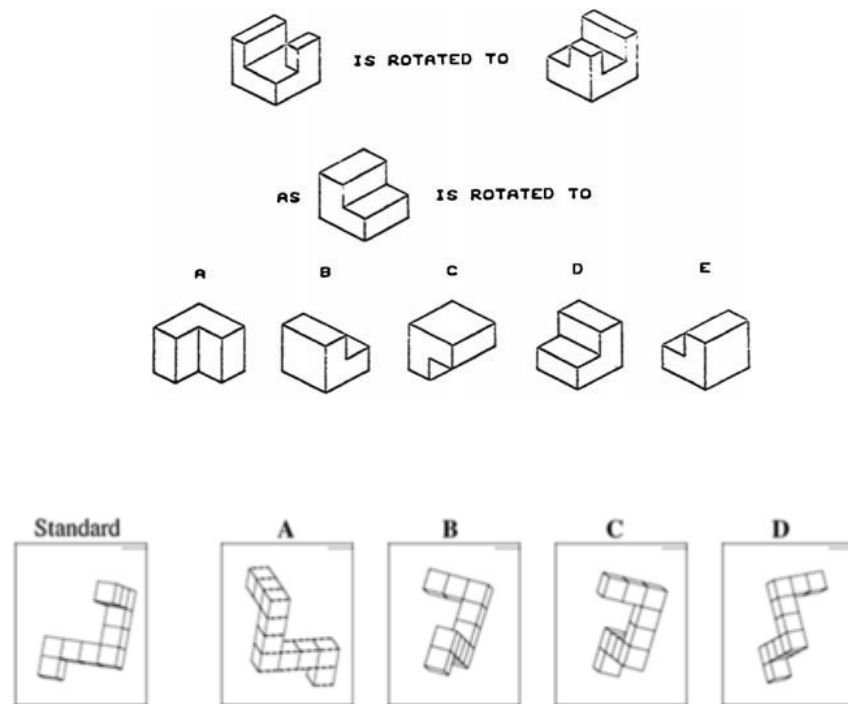


Figure 11: (top) PSVT:R, (Bottom) Vandenberg & Kuse's Mental Rotation Test

Vandenberg & Kuse's Mental Rotation Test (MRT) offers a standard view of an object, and four rotated views of this object: either a rotation of the original object or a rotation of the mirror image of that object. Test takers are required to distinguish between the mirror image and the standard image (an example is shown above in Figure 11). Much like the PSVT:R, the MRT has a long-standing use in research studies (Larrue et al., 2013). In some studies both of PSVT:R and MRT are used as measures of visualisation (Pontillo, 2010).

While the tests above were considered valid and reliable measures of visualisation, tests like the Form Board Test require two-dimensional visualisation (Mosoloane, 2013), whilst the MRT and Paper Folding (VZ-2) are more three-dimensional in nature. Sorby (2009) highlighted that Piaget theory of visualisation development placed 2D/3D distinctions within a hierarchal relationship. Specifically, that in the early stages of development,

between the ages of 3 and 5 years of age, most children had acquired the ability to deal with 2D puzzles that required them to select shapes that will fit together to make a target shape. Later in development, often by adolescence, children will be able to visualise 3D shapes, and be able to perceive how they will look if rotated or seen from another perspective; however, this task can still remain a challenge for some people, even into their late teens and beyond, if the object is unfamiliar to them. The final stage in Piaget's theory is when a person is able to combine 3D visualisation with measurement and is able to estimate volumes, distances, and mirrored reflections, for example.

From a design perspective, and particularly one based on graphic design, what the above operational measures of visualisation appear to avoid is colour, shade, or texture. Indeed, these fundamental visual qualities are uniformly absent from the examples above: whether 2-dimensional or 3-dimensional in task, they each offer black and white lines and shapes. Additionally, the visualisation of colour, for example, appears not to be a factor that is part of the theoretical frameworks and discussion. Such an absence is counter-intuitive in relation to the sort of training designers and artists receive both within higher education, and prior to beginning formal training in their chosen field of specialism. Yet, the prevalence of these tests in STEM subjects, and the historical value placed on the interpretation of schematics and technical drawings in engineering design may help explain why these measures were developed in their current format, and have continued relevance for some subject areas. However, for graphic design, are the measure of visualisation above a true reflection of the depth of mental representations that designers and artists use as part of their creative processes?

The work of Maria Kozhevnikov's et al. (2005, 2013) and Blajenkova et al. (2006, 2010)

would argue they do not. That what the measures above identified were only one of two distinct factors that underpin mental visualisation: the ability to visualise spatial relationships (such as locations, and rotations) and the ability to visualise objects (such as colour and shapes). Key to Kozhevnikov's et al. (2005) finding was that analysis of visualisation scores highlighted those who scored highly on spatial measure, performed poorly at object visualisation, and those who scored highly on object measures performed poorly on spatial visualisation. In a follow-up study, they identified that certain professions mapped onto these two groups, and, of particular interest for this thesis, that engineers tended to excel at spatial visualisation, and visual artists tended to excel at object visualisation.

In these studies, several of the measures discussed above were used. For example, spatial visualisation was operationalized through the use of the ETS Paper Folding Test, and a mental rotation task similar to the MRT. (Specifically, the task was a computerised version of Shepard and Metzler (1971) mental rotation task: a precursor to Vandenberg and Kuse's study and the basis upon which the MRT was developed). It used drawings of block shapes in various rotations and/or mirrored states, but participants were only shown one pair of blocks at a time (rather than several) and asked to decide if they are the same or mirrored. For object visualisation, Marks' (1972) Vividness of Visual Imagery Questionnaire (VVIQ) was used. This was a self-report instrument where participants were asked to rate how well they can visualise a number of scenes, for example: "The sun is rising above the horizon into a hazy sky". The VVIQ has a set of 16 questions rated on a 5 point scale. It has been used in several hundred studies and has been shown to correlate with objective indicators of visualisation. Kozhevnikov et al. (2005) also used two other measures of object visualisation. The Degraded Pictures task was based upon the ETS Snowy Pictures, and involved participants identifying a line drawing of a common

object (such as an umbrella) interwoven within a background of random visual clutter (representative of trying to decipher the shape of the object through a snow storm). The other measure was a Grain Resolution task. Participants were shown two words, and asked to decide which had a finer texture (granularity) than the other. For example, does a heap of salt have a finer texture than a heap of poppy seeds? The internal consistency for this task was a Cronbach's alpha level of .62, and it is likely that this moderate score is due to participants differing mental images of the objects they were asked to imagine rather than the vividness of their mental imagery. While, arguably, poppy seeds have a level of granular consistency, a person's visualisation of a heap of salt could range from the coarseness of rock salt to the fineness of commercial table salt: it is unclear in the study how this range of interpretation was counteracted during scoring for this task.

4.5.3.2 Suitable Measures of Visualisation

Given the range of options above, it was important for further study in graphic design that a battery of visualisation tests were required to measure both object visualisation and spatial visualisation. Object visualisation appeared well operationalized through the use of Mark's Vividness of Visual Imagery Questionnaire (VVIQ) and the Educational Testing Services' (ETS) Snowy Pictures test. Spatial visualisation through the use of the Purdue Spatial Visualization Test: Rotations (PSVT:R) and ETS's Paper folding test (VZ-2).

In terms of timescales, the ETS Paper folding test (VZ-2) and Snowy Picture test (CS-3) had a set time of 6 minutes each (3 minutes per part). 20 minutes was selected for the PSVT:R as this amount of time negated a gender bias that had been identified when shorter time conditions were used. VVIQ was an untimed test, and participants varied in the amount of time required to visualise the questions and respond: 5 to 10 minutes was estimated.

However, the total time required to undertake these four instruments was in the region of 56 minutes. Given that the written comprehension tests would take 30 minutes to complete, this was considered too much time for visualisation tests. Additionally, the use of time was skewed towards spatial visualisation. For this reason, the PSVT:R was prioritised over the ETS Paper Folding Test (VZ-2).

4.5.4 Limitations of These Findings.

The approach to mapping was built upon methods of analysis associated with textual coding undertaken in quantitative studies, particularly grounded theory. However, clear and direct methodological parallels, admittedly, are tenuous: in this study, an initial coding framework already existed in the form of CHC Theory, and analysis was based on the publicly available works of specific organisation rather than the private transcripts between a researcher and an informant. As mapping was undertaken by one researcher, it was not possible to check for agreement on frequency scores. With additional resources, and the option to collaborate with other researchers, the mapping process could be enhanced to improve the prioritisation of O*NET abilities, and how these connect to alternative models of graphic design competencies, like BTEC and NASAD.

4.6 Conclusions

The purpose of Study 2 was to map the conceptual connections between O*NET abilities and other graphic design competency models. Specifically, to explore how frameworks that underpinned O*NET (such as CHC Theory, and F-JAS) could be used to prioritise, and operationalize, graphic design competencies for experimental research.

Based on the findings above, the abilities of Visualisation (Vz), ranked 5th, and Written Comprehension, ranked 3rd, appeared important to BTEC, NASAD and O*NET's graphic design competency models as a whole. Through the identification of these terms, it

allowed existing research literature on F-JAS to suggest operational measures for research purposes. While F-JAS suggestions were limited for measures of visualisation, their identification supported the search for alternative measures.

Chapter 5, Study 3: Is the CAT a Reliable Method of Assessing Graphic Design Creativity?

5.1 Introduction to Study 3

The findings of Study 1 highlighted, given the relative novelty of the CAT to graphic design, there was a need to explore whether judges could show a suitable standard of consensus as has been achieved in other creativity research - coefficient alphas between 0.7 and 0.9 are a standard level of acceptable inter-rater reliability (Kaufman et al., 2008).

Prior to Study 1 few studies published in design journals had used the CAT but had shown sufficient levels of consensus within their respective domains. Combined with the longstanding use of the CAT in psychology based journals, the expectation was a graphic design CAT should follow a similar pattern of inter-rater reliability as other research. Clearly, there was a need to test this assumption, and this was the purpose of Study 3.

It is important to appreciate the context of this assumption on Study 3. In hindsight, given the findings of Study 3, 4 and 5, the basis for this assumption was misjudged. Study 3 was supposed to be a relatively straightforward research aim to identify an appropriate task to measure graphic design creativity using the CAT. However, the context of graphic design appeared highly sensitive to CAT research protocols, and a forensic approach to research design was required (as later used in Study 4 and 5). The outcome of Study 3 was that preconceptions of CAT protocols and expectations could not be taken for granted but, as will become clear, this was not the starting point.

5.1.1 A Task for a Graphic Design CAT

The task/design brief used for a CAT is a fundamental decision, one that can impact on

the validity and reliability of this measure (Kaufman et al., 2008). At the time of Study 3, a graphic design CAT had yet to be published, and such a choice on task would, understandably, be daunting for any creativity researcher. In contrast, as a sector, graphic design higher education had assessed creative potential for decades; setting specific design briefs to challenge and nurture students' creative potential was common. Whilst the practices of these two areas (the tasks/design briefs of graphic design education, and the application of creativity assessment through the CAT) had remained separate; it seemed a useful first step to bring these together.

In this respect, the ethos for Study 3 was explorative and open-ended, and for this reason, drew upon both quantitative and qualitative perspectives as the opportunities arose. Moreover, as an established academic within a design school, and director of a postgraduate creativity studies course, discussion with students and lecturing staff was a natural resource. By its nature, had a formal research approach been adopted it would have been at odds within these existing relationships. Hence, Study 3 was informal and cooperative. This enabled tacit knowledge to surface on what a suitable task for a graphic design CAT could be, as well as possible limitations and concerns. These conversations took place online through the MA Creative Thinking LinkedIn Group, personal emails, or in staff offices, or in design studios.

Two issues emerged that began to preference the selection of some graphic design tasks over others. The first issue was to identify a graphic design task that, theoretically, could be undertaken by a wide range of participants: a task that was neither overly skewed towards novices nor, conversely, towards intermediate or professional levels of graphic design ability. Second, the aim of Study 3 was primarily concerned with the level of consensus amongst judges on a graphic design CAT. Without the reliability of judges'

consensus, it was of little benefit if a task was suitable for a wide range of participants.

Task selection, for this reason, was initially speculative: premised on what might work, then test to see if it did. In a manner, parallels between the design thinking technique of fast/quick prototyping and Study 3 offer a useful context.

Where most CAT studies recruit two broad groups (one group of participants to create the outputs; one group of judges to assess the outputs) for Study 3, as highlighted above, it was the reliability of the judges on a given task that needed to be established first. A quick way to test this was to provide judges with existing outputs that applied to the task being explored. It was not required to have participants undertake tasks and generate new artwork if pre-existing artworks could be used.

Such an approach to CAT study is not without precedent. Baer, Kaufman & Gentile (2004), and Christiaans & Venselaar (2005) for example, each used work created under non-experimental conditions, and works that have been described as non-parallel creative works, i.e. “ones not created in response to the same prompts, but in response to very different assignments” (Kaufman et al., 2008, p.172). They were able to show that CAT inter-rater reliability could remain acceptable even when creative outputs are not generated under experimental conditions, or with the same prompts that generated the initial work.

In the sections below, two tasks are presented in detail, with their own relevant background literature, method, results and discussion, before a broad set of conclusions related to each is discussed. The rationale for this is that each task was tested for CAT reliability one at a time. Findings, in this way, were emergent, and the first task results

informed the research design of the second task. The first task was based on the seminal graphic design book *Watching Words Move* (WWM), the second the design of a T-shirt graphic. Each had historical and current usage in design education and appeared to give a wide range of participants the opportunity to express his or her graphic design creativity. Each, potentially, had a body of pre-existing artwork that could be made available for this research.

5.2 Watching Words Move Task

Within a graphic designer's education certain tasks/design briefs have stood the test of time, and are still used today on a variety of courses. Specifically, there is a task that has its basis in graphic design work from the 1960s. Its origin from a book called *Watching Words Move*, now considered to be a seminal work in graphic design and, as the quote suggests below, remains well regarded.

Words have the power to move. In 1962, a modest design studio created its own riff on that statement in the form of a small booklet of typographic brilliance, and changed forever how designers thought about the graphic potential of words. Decades later, the impact of *Watching Words Move* is still felt. Never before had the idea been so lucidly and playfully expressed that type itself could speak; that word-forms carried their own implied visual meanings; that the placement of letters on the page could suggest motion, narrative, emotion--just about anything. (Chronicle Books, 2009).

On the BA (Hons) Graphic Design course at The University of Central Lancashire (UCLan) this task has been part of the creative curriculum for the last few decades.

Equally, other graphic design courses and design courses follow similar design briefs. Watching Words Move, in this respect, appeared to be a task that could find wide acceptance from within the graphic design community.

As a task, it uses text only, and requires the participant to choose a word, and then visually communicate that word through the use of type. For example, the word "subtract" could become "Subtrct" or "-tract"; "add" could become "addd" or "+dd"; "stop" could become "s t o p!" (see Figure 12, Watching Words Move, p.27).

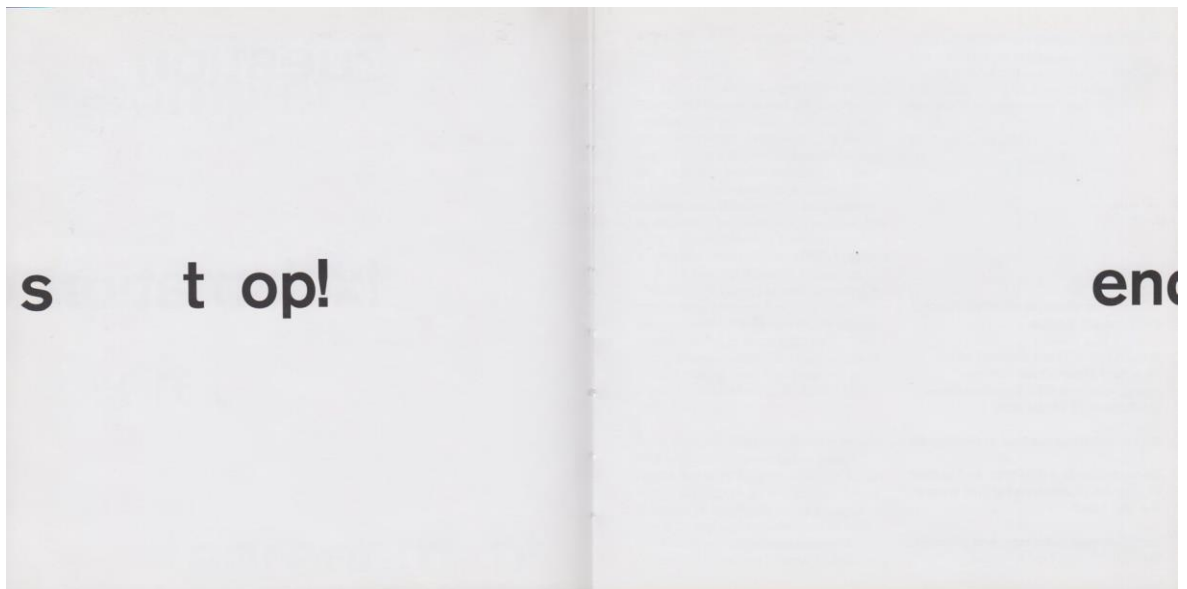


Figure 12: Example from Watching Words Move

5.2.1. Method

As described above, the approach taken with Study 3 was to be explorative and informal in the use of judges. The intention was to generate CAT data that could evaluate the WWM brief, but also offer opportunities for feedback from judges. Dialogue took place online through the MA Creative Thinking LinkedIn Group or personal emails, or in person at staff offices, or in design studios.

Following the precedent set by Baer, Kaufman & Gentile (2004) on the use of non-parallel creative works to test inter-rater reliability, it was unnecessary to recruit any participants to this study. What judges required was suitable stimuli to access their tacit knowledge on graphic design creativity; for this reason, plausible pre-existing artwork was gathered, and a suitable design brief developed that could be given to the judges.

This would be the first time the WWM task had been applied to the CAT, and it seemed prudent to check the design brief, instructions to judges and pre-existing artwork made sense. A simple and convenient way to achieve this was to ask students and alumni from the MA Creative Thinking to volunteer to trial this. Many of these students had studied the CAT as part of their course, and the practical experience of testing a CAT task prior to use by experts graphic design judges was beneficial for their development as creativity researchers. Moreover, this would allow CAT data to be gathered on novices' assessments which, when combined with experts' assessment, could offer further insight regarding CAT protocol and guidance.

5.2.1.1 Artwork Selection

The challenge with gathering a sample of works (much like the T-shirt Task to be discussed) was not straightforward. The graphic design staff at UCLan had archived student works, but on closer inspection, little was available for work that could be said to be pedestrian or mediocre; the works available were of exceptional quality. The same situation was found on the web. Indeed, most of the examples available were also either exceptional or highly creative. In essence, few examples of archived works of questionable quality could be found. These examples likely existed, but they were only available directly via current graphic design students or alumni, and this raised ethical implications for making such a request. This would be addressed in the subsequent study;

for this explorative study, however, enough examples could be gathered via the web to test the use of this task for CAT assessment. As will be discussed later in this chapter and through Study 4 and 5, the extent to which diversity of artwork may be crucially important for a graphic design CAT was not fully appreciated at the time.

An initial search of Google Image, using the phrase “watching words move” identified over 40 examples of pre-existing artwork. From this sample, 18 designs were chosen (shown in Figure 13 and 14) to serve as stimuli for the judges.

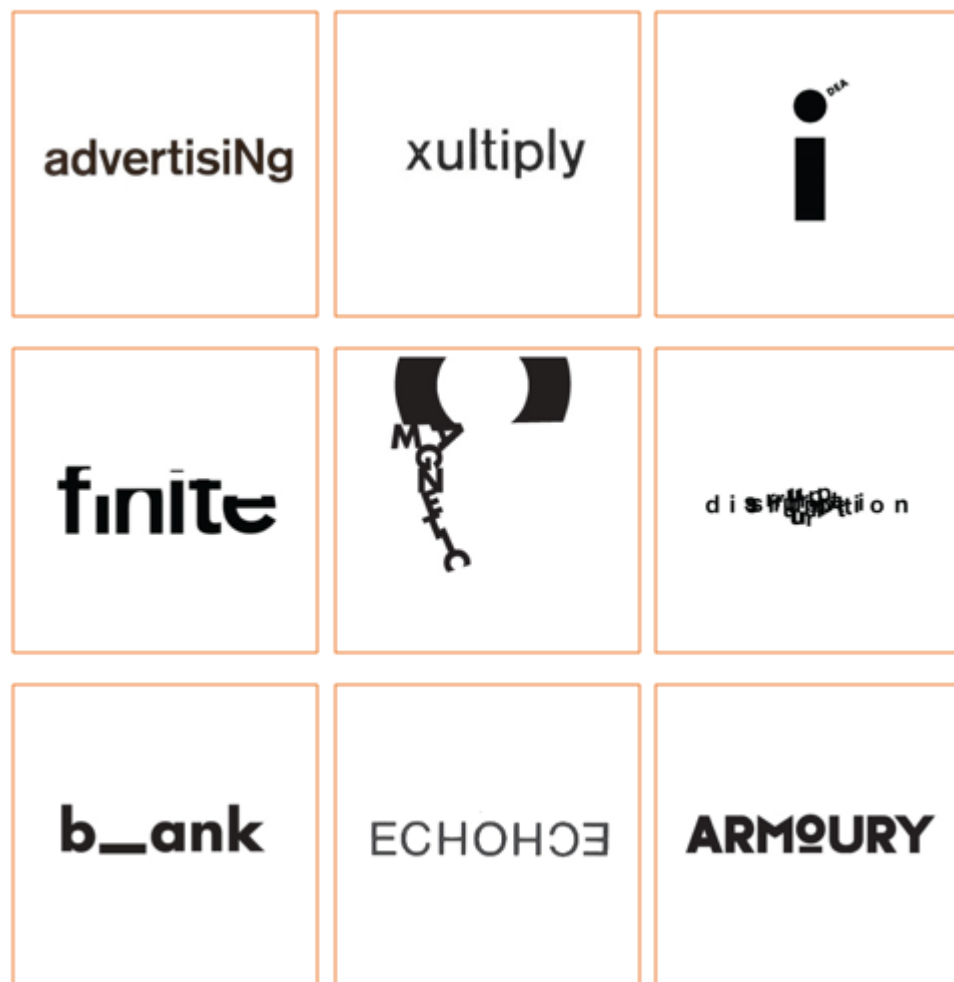


Figure 13: Artwork selected for the WWM task



Figure 14: Artwork selected for the WWM task (continued)

5.2.1.2 Instructions to Judges for the Watching Words Move Task

Separate to the practical issues of task selection was consideration of a number of decisions required for the adaption of CAT instructions to enable the rating of WWM artwork and the specific wording of those instructions for judges. For example, based on previous CAT studies, a poetry task presented to participants, and thus a substantial part of the instructions given to judges, would look as follows (Kaufman et al., 2008, p63):.

Table 20: Wording of a CAT task given to participants

On the next page, you are asked to write a poem using the format called SciFaiku. SciFaiku is a form of poetry derived from Haiku, a traditional Japanese poetry form composed of three lines of less than 17 syllables. The topic is science fiction. It strives for the directness of expression and beauty in its simplicity. SciFaiku also frequently strives for insightful commentary on the human condition. Here is an example:

on blackhole's edge
indecision
drifts me in

You can also write more than one stanza, following the same rule of three lines of each. Here is another example:

Hydroponics bay
a snail among stars
on the wide porthole glass.

Mid-spring, anticipating
the imminent cloning
of humans.

Bathing
her reptilian skin –
small bubbles on glossy green.

In the space provided below, please write a SciFaiku poem, with the theme of science fiction. You can write anything you like, as far as your poem follows the rules of Haiku (three lines of less than 17 syllables in one stanza). You should spend about 10 min on this, but please take your time

While the task above is for poetry, several defining features were abstracted to inform the wording of the Watching Words Move brief. Firstly, the tone of the instruction was relaxed; secondly, they provided an example of what creative solutions may be; thirdly, they established constraints and limitations within the task; fourthly, the amount of time given for the task was shown. Given this guidance, the following wording was developed (Table 21) for the Watching Words Move Task.

Table 21: Wording for WWM task given to participants

The brief given to the subjects was as follows:

Using black and white type only, turn the letters of a word into a graphic image that describes something about the word you have chosen. For example:

- The word "subtract" could become "Subtrct" or "-tract"
- The word "add" could become "addd" or "+dd"
- The word "stop" could become "s t o p!"

Within these constraints*, be as creative as you can be. You choose the word, the size, the type, and the page layout.

* The design must fit on one A4 sheet, and use black and white type only. You must complete this task within seven days of receiving this brief.

5.2.1.3 Developing the Rating Sheet for Judges

Equally, through adapting suggestions from Kaufman's et al. (2008) *Essentials of Creativity* and guidance on the use of the Consensual Assessment Technique, the following instructions were developed (Table 22):

Table 22: Instructions give to judges

Please look through these graphics twice. The first time, assign a Low, Medium, or Higher rating. The second time, assign a numerical rating between 1 to 6, with 1 being the least creative and 6 being the most creative. There should be a roughly even number of graphics at each of the six levels but the numbers needn't be exactly the same. It is very important that you use the full 1-6 scale, however, and not assign almost all graphics the same rating.

There is no need to explain or defend your ratings in any way; we ask only that you use your own sense of which are more or less creative. Simply write the number on the paper below (1, 2, 3, 4, 5, or 6 -or, if you would find it helpful, any decimal from 1.00 to 6.00 -but nothing below 1.00 or above 6.00 please).

As judges needed to have a means to record their assessments, a rating sheet was given with the instructions (Table 23).

Table 23: Rating sheet provided to judges

	Low		Medium		High	
	1	2	3	4	5	6
advertising						
multiply						
idea						
finite						
magnetic						
disruption						
blank						
echo						
armoury						
savory						
copy						
superstitious						
missing						
sky-high						
soul						
top						
splitting						
vision						

5.2.1.4 Participant Judges

Ten judges rated the 18 artworks. Five would be considered to be a novice or intermediate judges (i.e. they had no experience of graphic design, or had some experience but would not be considered experts). In addition, five judges took part that would be considered experts in graphic design. Expert judges either worked as full-time academic staff teaching graphic design or worked full time as professional graphic designers. Each judge was given the assessment sheet above, and jpeg images of the 18 designs (fig. 13 & 14).

5.2.2 Results

SPSS was used to compute Cronbach's alpha (the level of inter-rater reliability). The alpha for the novice/intermediate judges was 0.81; the alpha for the expert judges was 0.69. The illustration below (Figure 15.) compares the expert judges' ratings relative to the

novice/intermediate judges' ratings for the same artwork.

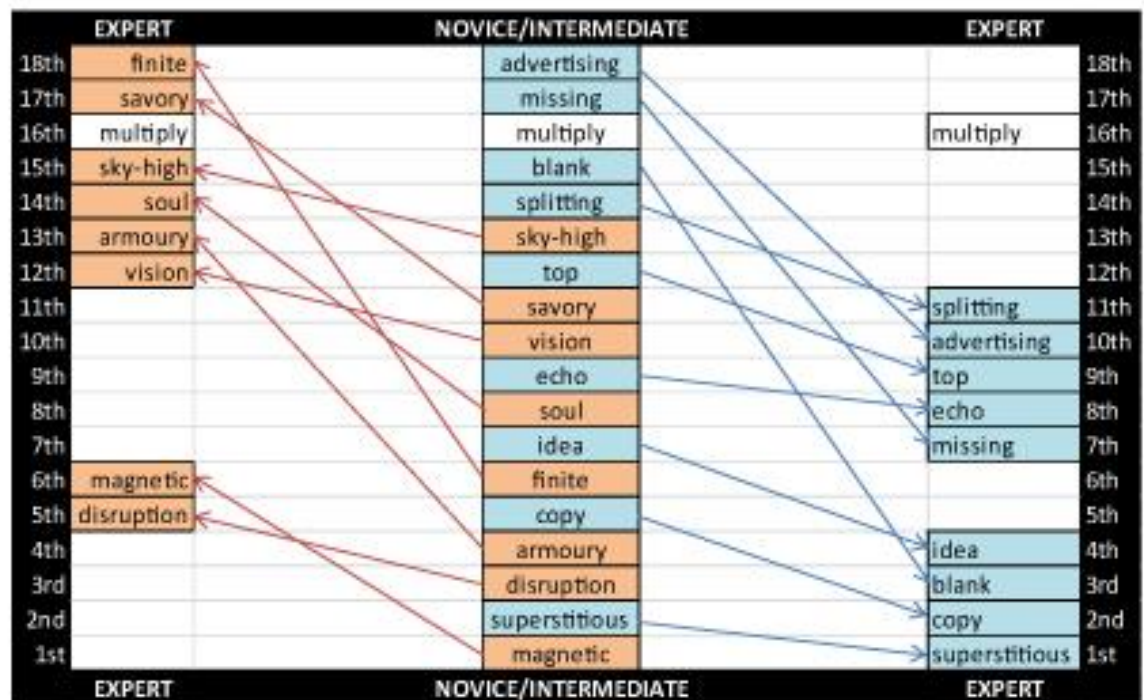


Figure 15: Comparison of expert scores with novice/intermediate scores for the WWM task

Arrows on left show where expert judges' rated a work lower than novice/intermediate judges; arrows on right show where expert judges' rated work higher than novice/intermediate judges. The item "Multiply" ranked 16th for both groups of judges.

With inter-rater reliability alphas between 0.69 and 0.81, it was interesting to note the lower level of consensus experts achieved relative to novices/intermediates for this task. Both alphas are acceptable, but the aggregated rankings were quite different: dramatically so in some instances. For example, one of the designs was the word "advertisiNg"; this artwork was unanimously rated the lowest by the novice/intermediate judges; expert judges rated it at mid-level (10th out of 18th). An expert judge highlighted that advertising was about making things "stand out", thus the way the capitalised "N" stands out in this graphic was reasonably clever in making that point. For the novice/intermediate judges none of them understood the reason for why the "N" was capitalised, and indeed most commented on this as a detrimental feature.

5.2.3 Discussion of WWM Task Results

In the "advertisiNg" example, it would appear domain knowledge influenced creativity rating. Overall, this finding is not new; it confirms other studies that highlight the importance of using domain experts for CAT tasks, and are discussed further in section 5.4.2 Suitable Judges. However, the alpha for the novice/intermediate judges was not lower than the experts; indeed it was considerably higher.

It may be the case that the high level of consensus from the novice/intermediate group was influenced by domain independent knowledge of words rather than domain dependent knowledge of graphic design. For the graphic design experts, graphic design creativity is a construct that is made up of both linguistic and visual creativity, (for example type and image) and the two can be inseparably interwoven within some artworks. To what extent ability with words and text is correlated with graphic design creativity is further explored in Study 5.

From a different perspective, prior to undertaking this assessment the novice/intermediate judges engaged in debating the merits of the Watching Words Move Task in relation to creativity assessment. This took place in a private group on LinkedIn. Relevant concerns highlighted the need for professional graphic design software to create a design, versus presenting the designs as hand-drawn concept sketches (an observation that re-occurs later in this chapter and is discussed further).

As a final reflection on this study, the design of the rating sheet in itself became a point of consideration. Whilst the rating sheet was for the most part purely a means to record a judge's responses; its design also had the potential to visually reinforce the instructions to judges or, conversely, obscure them. A key feature of the instructions to judges was to use the whole rating scale from one to six, specifically:

There should be a roughly even number of graphics at each of the six levels but the numbers needn't be exactly the same. It is very important that you use the full 1-6 scale, however, and not assign almost all graphics the same rating.

Yet, the example below from an expert judge (Figure 16) seemed to overlook this instruction. Predominately, the majority of ratings are either 1 or 2, with the scales 3 and 6 not used at all.

		Low		Medium		High	
		1	2	3	4	5	6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	advertising	✓					
	multiply	✓					
	idea				✓		
	finite	✓					
	magnetic		✓				
	disruption	✓					
	blank					✓	
	echo	✓					
	armoury		✓				
	savory	✓					
	copy		✓				
	superstitious				✓		
	missing		✓				
	sky-high	✓					
	soul	✓					
	top	✓					
splitting	✓						
vision		✓					

Figure 16: Example of restricted rating sheet

Most judges in the study used the full scale, but this example highlighted a weakness in the current rating scale: that for experts from a visual culture, there were no visual cues to complement the text-based instructions. For this reason, this was a point directly

addressed in the T-shirt Task.

In summary, the Watching Work Move task was reliable, though the task could be blurring linguistic and visual communication creativity, and this may mean it is skewed towards suggesting a relationship between linguistic skills and graphic design creativity by measuring both of these constructs, rather than visual communication creativity in isolation. This was something to test further, and to be cautious that the task chosen to measure creativity could infer relationships between skills/abilities or knowledge that may or may not exist. The broader point considered was that for testing the reliability of a suitable CAT task for graphic design creativity, alpha level was only one part of reliability; rankings also appeared to be important, and the results of this study highlighted the need to use domain experts in future graphic design CAT studies. Lastly, it was considered useful to develop a rating sheet that would visually reinforce the use of the whole rating scale for judges.

5.3 T-shirt Design Task

For several years the Open University has undertaken an introduction to design thinking, of which hundreds of students each year have enrolled. As part of the assessment for this module, students have been briefed to design a graphic which could be transferred to a plain white T-shirt. The graphical image was required to make reference to their hands, and students were asked to explore three different approaches in relation to their hands, and then selected one of these routes as the basis for their final T-shirt design.

The first route was to look at the features of their hand from an aesthetic and objective perspective. The second route was to think about a story that relates to a feature of their hand, for example, how they got a scar, or a story about some hand related skill, such as

a sport or playing an instrument. The third route was to explore how they use their hands to communicate through gestures.

Having thought about each of these routes, they were required to take a photograph that expressed their ideas and uses these photographs as a basis for a graphic concept for their T-shirt designs. For the concept that involved the aesthetic features of their hands, they were asked to trace around the photograph and make a simple black-and-white line drawing and then to alter and manipulate this shape using two principles: adjusting the scale/proportion, or trying to achieve balance/symmetry through composition and layout. With the story about their hand concept, they were asked to draw around this image and then fill in the object that the hand is holding. With the gesture concept, they were to draw around the image and then use text to describe this gesture, placing this text at some point in the composition. Having done this, they chose what they considered their best design, and transferred this design to a plain white T-shirt.

5.3.1 Method

Given the findings that emerged from the WWM data, the method for studying the T-shirt task both retained and changed as a result. Like the WWM brief, the approach was explorative and informal in the use of judges. The intention was to generate CAT data that could evaluate the T-shirt task, but also offer opportunities for feedback from judges (which took place within staff offices or in design studios). Similarly, the focus was upon offering judges suitable stimuli to access their tacit knowledge of graphic design creativity through pre-existing artworks and a suitable design brief. Where the T-shirt tasks differed (as detailed further in this section) was in the use of graphic design experts only; increase in the number of judges from five to eight; changes to the rating sheet given to judges, and access to artwork from a complete student cohort rather than from Google Images. At

the time of this study, research governance permitted the use of this artwork for explorative research purposes within the University. With the public version of this thesis, subsequent guidance from the Open University Human Research Ethics Committee clarified the T-shirt artwork should be withheld from publication. Hence, no images of the artwork provided to judges are shown.

5.3.1.1 Artwork Selection

For the context of this study, the decision was taken to use a sample of designs to explore the reliability of the CAT. Hundreds of designs were available. However, this was too many examples for this study; indeed, around 20 designs would be adequate. The issue became one of selection.

For confidentiality reasons, those photographs which showed a person's face were excluded: this roughly accounted for discarding a third of the designs. The remaining images could be described as a photograph of the T-shirt on a hanger or laid out on a surface, a person modelling the T-shirt but only being able to see their torso in the photograph, or flat artwork showing the design.

The rationale was to select on the basis of flat artwork only. The benefit from this was to minimise the number of biases that may impact on the judges' assessment, for example, the gender of the model, the aesthetics of the person modelling the T-shirt, whether the T-shirt was ironed or not, the quality of the photograph in relation to lighting and such. None of these factors were issues with the flat artwork. With the flat artwork, however, a number of presentation issues were highlighted that could impact on the judges' assessment, but these were of a graphic design nature, for example, technical drawing ability, thus, were seen to be appropriate to this study. From designs left, 18 were selected, and each was represented onto an A4 sheet (any details that could identify individuals, such as a name,

or a photograph of someone modelling the T-shirt, were removed).

5.3.1.2 Instructions to Judges for the T-shirt Task

Each judge was given the instructions shown in Table 24. A point to note in this instruction was the amendment of wording related to the full use of the rating scale. In the Watching Words Move instruction the wording stated:

There should be a roughly even number of graphics at each of the six levels but the numbers needn't be exactly the same. It is very important that you use the full 1-6 scale, however, and not assign almost all graphics the same rating.

For the T-shirt Task this was edited to:

There should be an even number of graphics at each of the six levels. It is very important that you use the full 1-6 scale, however, and not assign almost all graphics the same rating.

The option to have an "uneven" number of ratings remained, but this is much more implied in this wording, rather than the explicit phrase used previously. Additionally, this aspect was reinforced through a new rating sheet.

Table 24: T-shirt instructions given to judges

Background to this task

In xxxx, xxx students undertook an introduction to design module at a UK University. As part of the assessment for this module, students were asked to design a graphic which could be transferred to a plain white T-shirt. The graphical image was required to make reference to their hands, and students were asked to explore one of three different approaches:

- to look at the details of their hands from an aesthetic perspective;
- to think about a story that relates to a feature of their hands (for example, how they got a scar, or a story about some hand related skill);
- to explore how they use their hands to communicate through gestures.

How to rate these graphics

Please look through these graphics twice. The first time, assign a Low, Medium, or Higher rating. The second time, assign a numerical rating between 1 to 6, with 1 being the least creative and 6 being the most creative. There should be an even number of graphics at each of the six levels. It is very important that you use the full 1-6 scale, however, and not assign almost all graphics the same rating.

There is no need to explain or defend your ratings in any way; we ask only that you use your own sense of which are more or less creative. Simply drag and drop a graphic onto one of the numbered boxes in the PowerPoint (file attached labelled: Judges score sheet), and save this as a new PowerPoint file.

To view the detail of each graphic a further PowerPoint is attached (labelled: T-shirt graphics). After opening this file, if you press F5 on your keyboard, you will be able to view these graphics as a slide presentation.

5.3.1.3 Development of rating sheet

With the T-shirt Task, the lack of a clear title for each work made the previous rating sheet difficult to use. For internal purposes, labels from 1 to 18, or A to R, could be used for identification, but this was felt to be too abstract and awkward for judges to use. A PowerPoint file was developed that allowed judges to drag and drop an image onto a rating sheet (Figure 17). It was also possible to use a paper based version, where judges could place a small black and white image on the rating sheet for their assessment.

The main consideration for the rating sheet, however, was to reinforce the use of the whole rating scale graphically. Thus, for the 18 artworks, there are 18 identified places, each one with a corresponding rating.

Low		Medium		High	
1	2	3	4	5	6
1	2	3	4	5	6
1	2	3	4	5	6

Figure 17: A PowerPoint drag and drop rating sheet was used for the T-shirt Task

5.3.1.4 Participant Judges

8 judges took part, all of whom would be considered graphic design experts. Four were full-time academics teaching on graphic design courses, and three were full-time graphic design practitioners. The remaining judge worked between graphic design practice and teaching of a graphic design course. Each judge was given the assessment sheet above and a PowerPoint with the 18 images. Each judge independently rated the 18 artworks, after reading the instructions

5.3.2 Results

SPSS was used to compute Cronbach's alpha, and the alpha was 0.56. This is below the 0.7 levels of consensus required for reliability. Indeed, from the perspective of previous CAT studies, with regard to reliability, this level of alpha was a cause for concern. The reason for the low alpha score was unclear, but several possibilities are discussed below. For example, did the method and procedures contribute to this level of consensus: was the T-shirt Task too complex; the drag and drop rating sheet method too restrictive, or was some form of sampling bias through the images selected at play, or in the selection of

judges? What was clear was this level of alpha suggested these judges, rating the T-shirt Task, in the way they did, with the artworks available, did not concur sufficiently on their ratings of creative output in graphic design.

5.3.3 Discussion T-shirt Task Results

Why was the alpha level low for the T-shirt Task? In preparation for observations regarding technical execution, it is worth stating that for some graphic design projects (but this could equally apply to design more broadly) the transition from a concept to a final output can be fraught. Conceptual ideas that are novel at the outset are, for a variety of reasons, unable to be usefully developed, and the final output does not represent the potential for creativity it first suggested.

For certain researchers, assessing conceptual creativity is a more accurate reflection of creative output in a domain; they argue for creative concept being the key indicator for measuring high creative ability (Cowdroy & de Graaff, 2005). Technical execution or, for want of a better label, the "crafting" required to take a concept towards a final output, is secondary to the primacy of creating the concept.

After rating the T-shirt artworks, some judges clearly expressed that they were rating artworks from a conceptual perspective (with the focus on the creativity of the idea, regardless of the level of technical execution). Other judges may not have done so. Though never stated by expert judges in this study, it is feasible that, conversely, a judge may value the technical execution of a piece of graphic design over the conceptual idea. If either of these scenarios are probable, then the use of the CAT to measure graphic design creativity may need refinement to be optimised for this domain. Perhaps technical execution is prevalent within graphic design creativity when judged by professional graphic designers, and more explicit instructions are required for this domain?

It must be highlighted, however, that this issue of technical execution versus creativity is not new, indeed, this was a factor directly considered in Amabile's early CAT studies (1982, 1996), the outcome of which was that technical execution was distinct from creativity, and thus the CAT was shown to be rating creativity, not technical execution. The findings were also future proofed at the time by the acknowledgement that this clarity of distinction may not apply to all domains. Given the low alpha for the T-shirt Task, perhaps one of those domains could be graphic design?

A judge's preference, or not, for technical execution relative to conceptual creativity was also found in other post assessment comments. Some graphic design experts explained they would consider visually resolved outputs (i.e. those that were "Mac'd up") more harshly as a result of their appearance to being finished outputs. Such potential for bias with "Mac'd up" outputs (ideas presented via some form of computer software: such as Adobe InDesign, Illustrator, etc.) relative to hand drawn concepts, may again link to issues of technical execution versus the creativity of the concept. A further possibility is that some judges may have considered all work they were shown needed to be "Mac'd up", and those that were not were rated lower despite their conceptual merit.

Whether a hand-drawn concept sketch is blurring associations between defining creativity as an output or starting to move towards a creative process, could be theoretically ambiguous. What can be argued is that a hand-drawn version for a CAT task is testing the conceptual creativity of a graphic design, whereas the "Mac'd up" version is testing both conceptual creativity and begins to imply something of the creativity required to nurture a concept through to a finished output. The implications here is that for the T-shirt Task,

which contained a mixture of hand-drawn and Mac'd up artwork, this difference in format could be a factor in the low level of consensus among the judges.

A further consideration was whether the rating sheet used for the T-shirt Task was too restrictive in visually reinforcing instructions to use the whole rating scale: effective to the point that judges rated only within the boxes allowed. In this respect, the situation was opposite to the rating sheet used for the Watching Words Move Task, and could be forcing similarity of ratings that would not exist without the rating sheet boxes.

A final observation was that some expert judges suggested a lack of consensus could be reflective of a diversity of opinion on graphic design creativity amongst these judges (many of whom worked with each other, and where diversity of opinion amongst team members is an asset). That is an interesting possibility, and perhaps is a point that reinforces the benefits of only selecting one person per design studio. It is, also, an observation that does not fit neatly within the theoretical perspective of the CAT. Domain expertise is supposed to confer a tacit, if not explicit, shared understanding of creativity within the domain, rather than nurture polarity.

5.4 Broad Discussion

From these two studies, the main question to consider was why the inter-rater reliability of expert judges was below acceptable levels? Reliability measures of 0.56 for the T-shirt Task and even 0.69 for the WWM task were not ideal. Was something happening in graphic design that warranted caveats for research design, or was this domain challenging the assumptions upon which the CAT is based (i.e. domain experts can independently agree on creativity to an acceptable level of inter-rater agreement)?

The discussion below suggests a number of caveats in order to optimise the CAT method as a measure of graphic design creativity: specifically, the influence of technical execution on ratings of graphic design creativity; the background of judges; the range of artwork available, and the suitability of a task for research purposes.

5.4.1 Technical Execution

Artwork in graphic design can vary in its level of technical refinement. At one end of the spectrum are conceptual artwork, where the seed of an idea can be perceived: even if an artwork lacks refinement in, for example, font selection, layout or composition, the creativity of the idea can still be evaluated. At the other end of a technical spectrum is finished artwork; artwork that is ready to go to print or publication, where every aspect of visual communication has been crafted and refined to perfection by the designer.

The studies highlighted that judges' preferences for technical execution differed with regard to judging graphic design creativity. For example, their preference for technical quality appeared to be heightened or subdued relative to the medium of the artwork (hand-drawn sketches, artwork created with the use of computer software, or a combination of both mediums).

Within the CAT literature, some researchers have created instructions that directly address this issue of discounting technical execution (Baer, 1993). Other researchers (the majority of CAT studies) have not done so with little adverse impact: some ask judges to rate creativity alongside technical execution and aesthetic appeal (Amabile, 1982; Christiaans & Venselaar, 2005; Valgeirsdottir, Onarheim & Gabrielsen, 2015), some only do this the first time they undertake a new CAT task (Hennessey, 1994; Kaufman, Plucker & Baer, 2008); some do not. Thus, the issue of technical execution ranges from explicit,

through to implied, implied only once, or not mentioned at all.

Given this range of approaches, if graphic design experts are not specifically guided to discount technical execution then ambiguous instructions may impact on the level of consensus: could differences in technical preference explain low inter-rater agreement in previous studies? They could, and the suggestion was this aspect needed to be explored through further study.

5.4.2 Suitable Judges

The debate over the use of novice or intermediate judges is a contentious one for the CAT. Due to the relative ease to which the general public or students can be recruited compared to domain experts, there are clear advantages if CAT researchers do not need expert judges. However, the evidence base for this is mixed, as are the findings here from Study 3.

Broad guidance on the use of the CAT advocates that judges are required to be experts in the domain to which the task is associated, and that “The validity of the CAT is grounded in the fact that experts in a domain are the final arbiters of what is creative (or otherwise valued) in a domain.” (Kaufman, Baer, Cole & Sexton, 2008, p.175). Indeed, Kaufman, Plucker and Baer (2008) suggest that: “Judges should have a level of expertise that is clearly higher than the presumed level of expertise of the subjects creating the artifacts.” (p.74). Such guidance is based on previous CAT studies that have set out to explore if it is appropriate to replace expert with non-experts and find in comparison to experts raters, non-experts do not yield a sufficient level of consensus. Furthermore, where consensus levels are improved by increasing the number of non-expert raters (n = 106) the correlation between expert and non-experts total ratings is not similar enough to replace

expert raters.

In study 3, whilst novice/intermediate judges - relative to experts- achieved higher levels of consensus for the WWM task (0.81 compared to 0.69), the rankings for artworks did not match that of the experts (as can be seen in Fig. 15). For this reason, the suggestion here is to continue to use domain experts only, as other CAT researchers have advocated.

Within the context of graphic design, a further point is that domain experts can be based within professional practice (i.e. full-time graphic designer), or educational practice (i.e. full-time graphic design lecturers teaching on undergraduate and postgraduate graphic design courses). Previous research in other design domains (Jeffries, 2009) has suggested that the values shared on creativity between academics and practitioners are not as polarised as populist views can imply. However, whether this is the case for graphic design is unclear, and thus a cautious approach to CAT studies, in this domain, would be to gather experts from either academia or professional practice, but not to use both within the same group (that is until research findings can show otherwise)

5.4.3 Range of Artwork

It may be the case that the artwork sampled for the two studies lacked diversity. The likelihood is that the artworks gathered for the WWM task were mostly gathered from highly accomplished graphic designers. From the sampling of T-shirt artworks, the work may have, for the most part, clustered to low or medium quality (purely through sampling bias, and the introductory nature of the module these students were undertaking).

Whilst the CAT is a method that highlights ratings be relative to other works within a sample, it is feasible that artworks too similar in quality pose a more challenging task for judges than those that show more diversity.

Such a consideration is contentious. Several CAT studies highlight the real world basis of the technique, and with acceptable inter-rater reliability, there is little reason to question further. However, most CAT studies are research studies, and whilst parallels can be drawn between CAT methods and those used by judges of, for example, professional competitions or awards, the independence of judges does not happen throughout the rating process as it does in the CAT. Much like other consensual techniques, such as the Delphi technique, competition judges do tend to confer with each other towards the end of the process. A judge's initial assessment may be independent of the panel, but towards the final stages of evaluation, debate and compromise are not uncommon in order for consensus to be achieved. This does not discredit CAT methodology, but it does suggest that the real world assessment of highly creative artworks may require more debate amongst judges than the CAT allows.

If this is the case for highly creative artworks, then why not for other skewed samples, be they predominantly low or medium in quality. It is beyond the scope of this research to explore this in greater detail, but the argument here, given the findings, is that some form of pre-test for diversity of artwork, or stratified sampling of participants, may be useful.

5.4.4 T-shirt Task

As a final consideration, the T-shirt Task may be too complicated for the purposes of this research.

Firstly, as a task, it was designed for an educational purpose with a broad scope. For example, it was testing the development of a student's design and creative process, and one that took place over an extended period of time, that is, extended relative to the more experimental tasks used by many CAT studies. Additionally, two expert judges also

commented on the ambiguity of the T-shirt brief relative to the Watching Words Move brief: it may be the task itself could explain the low level of consensus for the T-shirt ratings for expert judges.

Secondly, as discussed previously, the artwork ranged from hand-drawn sketches to "Mac'd" up. This difference in format could suggest associations a judge may have with "Mac'd up" outputs (ideas presented via some form of computer software: such as Adobe InDesign, Illustrator, etc.) relative to hand drawn concepts, and this may again link to the issue of technical execution versus the creativity of the concept. After judging, some judges clearly expressed looking at all works more from a conceptual perspective (with the focus on the creativity of the idea, regardless of the level of resolution). Others suggested they would consider visually resolved outputs (i.e. those that were "Mac'd up") more harshly as a result of their appearance to being finished outputs. A further possibility is that some judges may have considered all work they were shown needed to be technically resolved, and those that were not were rated lower despite their conceptual merit. These types of distinctions appear important and could explain the low alpha for the T-shirt Task

Thirdly, some artworks incorporated text within the T-shirt graphic, whilst other did not. In contrast, the WWM task used design elements that are predominately based on text. The relationship between image and type is an interesting one, and a case can be argued that one of the distinguishing features of graphic design is its use of both forms of visual communication, and the creativity to turning words into icons, or use of images to say *a thousand words*. However, the combination of type and image in one task may add a layer of complexity, that is, relative to a type only task, or an image only task. From a research perspective, the isolation of type from the image may have benefits. It could be possible

that a person is creative with type but not with image, or vice versa, and by having a Type Task and an Image Task could give some insight into this. Equally, by combining a Type Task CAT score with an Image Task CAT score could offer a fruitful measure of graphic design creativity in itself; something which would not be possible if both these design elements are integrated within a single task.

Fourthly, the rating sheet used for the T-shirt Task was directly developed to reinforce instruction to use the whole rating scale. All of the judges used the full rating scale, but none took up the option in the instruction to have more than three items under a specific rating. In this way, the design of the rating sheet itself could have skewed the level of consensus, with its designation of boxed areas. For this reason, it was advisable to look at a more flexible approach to the rating sheet design, one that reinforced the use of the whole scale, but not to the extent of overly stipulating the number of items under a specific rating.

5.5 Conclusions

While, at the outset, the CAT was supposed to work for graphic design as it had in other design domains, the findings of Study 3 highlighted this was not the case in practice. Each of the issues discussed above may have contributed to the low level of consensus achieved in the Study 3, and the influence of technical execution on ratings of graphic design creativity appeared to be a significant consideration. Equally, a graphic design CAT appeared more sensitive to research method than had been assumed. Given this, the suggestion for future study was for artworks to contain image only or text only; that artwork would be hand-drawn or Mac'd up (but not contain a combination of both); that rating sheet design needed to be reconsidered, and that the task should be able to be completed in a relatively short time scale, or at least one comparable to the WWM task. By

refining these issue of method and research design in future studies, it would be possible to optimise CAT protocols for graphic design creativity, and this was the purpose of Study 4 and chapter 6.

Chapter 6, Study 4: Optimising the CAT to Measure Graphic Design Creativity

6.1 Introduction to Study 4

This chapter builds on the findings of Study 3, and by extension, further tests the reliability of the CAT as a measure of graphic design creativity. As discussed in Study 1 and 3, few research studies had applied the Consensual Assessment Technique (CAT) as a measure of graphic design creativity. Indeed, how to interpret the CAT research method, and its translation from experimental psychology to the *real world* practice of graphic design creativity appeared to require a number of caveats not immediately apparent from previous research. One of the suggestions based on the results of Study 3 was inter-rater reliability may be unduly influenced by a judge's preference for technical execution. Study 4 extended and tested this hypothesis, and line of enquiry. It compared the influence of CAT instructions to discount technical execution from creativity ratings against those that make no such stipulation. Additionally, it sought to explore the level of inter-rater reliability for an image based graphic design task and a text-based graphic design task. By isolating creativity with type from creativity with image it enabled a degree of experimental control over these two factors; which in Study 3, when combined within one task, as with the T-shirt task, may have contributed to the low level of inter-rater reliability.

6.2 Literature Specific to this Chapter

Since the systematic review of Study 1 in 2012, the popularity of the CAT continued to grow within design research, and two years later a PhD study applied the CAT to traditional graphic design (Wojtczuk, 2014).

While acknowledging the previous discussion on Silvia et al. (2008) and Dineen & Niu (2008) - see section 3.5.4 (Use of CAT in graphic design) Wojtczuk's research applied the

CAT to graphic design through the use of professional practitioners of graphic design (in contrast to design educators), and through a design brief directly related to graphic design.

The task was to design a poster for “a specific event concerning ecological packaging” (p.100), and Figure 18 shows two example of the 21 posters gathered for the study. Each poster was designed by a single student, over two months, as part of their graphic design degree at a French graphic design and visual communication school.



Figure 18: Two examples of poster design used

A number of findings on the influence of judges' background on design creativity assessment were highlighted by Wojtczuk, and of particular importance was the low level of consensus achieved by professional graphic designers' ratings of graphic design

creativity for the poster task: an alpha of 0.42. This was a particularly interesting finding, as Wojtczuk's research was also replicating the challenges encountered in this thesis, through Study 3, but in a different cultural context and with a different graphic design task.

Fieldwork for the Watching Words Move Task and T-shirt Task (presented in chapter 5) was taking place at a similar time to Wojtczuk (around 2011, 2012), and as discussed in chapter 5, was also finding low levels of consensus among professional graphic design judges when using the CAT: 0.56 for the T-shirt Task and even 0.69 for the WWM task was not ideal.

in order to optimise the CAT method as a measure of graphic design creativity Study 3 highlighted the influence of technical execution on ratings of graphic design creativity; the background of judges; the range of artwork available, and the needed for an image only or text only task. Each of these points has been covered in greater detail in chapter 5.

Based upon such caveats, the following study was specifically designed to address the following questions:

- Can professional graphic designers achieve inter-rater reliability at or above 0.7 for an image based graphic design task and a text-based graphic design task?
- If they can, do CAT instructions to discount technical execution increase inter-rater reliability when compared to instructions that make no such stipulation?

6.3 Method

6.3.1 Broad Research Design

As is frequently the custom for CAT studies, two broad groups of participants were

required: those who would generate the creative outputs (participants), and those who would assess the creative outputs (judges). For many CAT studies outside design, participants are recruited and undertake the creation of a piece of work under experimental conditions -this is because the purpose of such studies is often to test the influence of teaching or environmental factors that may impact upon creativity.

In this study, as in Study 3, it is the judges, rather than the participants that are the focus. Thus, the method follows a research direction set by, for example, Baer, Kaufman & Gentile (2004), and Christiaans & Venselaar (2005) for the use of work created under non-experimental conditions. It has been shown that CAT inter-rater reliability could remain acceptable even when creative outputs are not generated under experimental conditions.

Indeed, for Baer, Kaufman & Gentile, the creative outputs assessed in their 2004 paper, had been created without consideration that they would become part of a research study. Similarly, the graphic design artwork in Study 4 was created as a natural result of engagement with a university degree. Specifically, the study gained consent from participants to use type only and image only artwork created during week two and three of the BA (Hons) Graphic Design course at UCLan. These two tasks provided 30 type and 30 image examples to be independently assessed by 16 professional graphic designers using the CAT.

The dependent variable was the instructions given to judges, and each judge was randomly assigned to receive different instructions for each task (an Image Task or Type Task). Using a 2x2 experimental research design (task x instructions), professional graphic designers were randomly assigned to receive instructions that emphasised

discounting technical execution from creativity ratings, or receive instructions that gave no stipulation on technical execution. Data was analysed for inter-rater reliability, and appropriate statistical analysis used to compare the influence of different instructions: the purpose being to consider if different CAT instructions had a significant increase on consensus for either image only or type only artwork.

6.3.2 Ethical Considerations

As this project was between two universities, authorisation was sought, and obtained, from ethics committees at each. The guidelines followed in this study were those of the British Psychological Society (BPS): specifically the 2009 Code of Ethics and Conduct, and the 2010 Code of Human Research Ethics. In addition, the Open University Code Of Practice For Research and Those Conducting Research, the Open University Ethics Principles for Research involving Human Participants, the University of Central Lancashire's Code of Conduct for Research and the Ethical Principles for Teaching, Research, Knowledge Transfer, Consultancy and Related Activities.

The project adhered to the Data Protection Act 1998, and more specifically the data protection codes of practice and policies for research, as set by the Open University, and the University of Central Lancashire. Names, contact details, and anonymous codes were password protected on an Excel spreadsheet: only the principle investigator knew the password (an 8 digit password of letters and numbers). This digital document was only accessible via a dedicated encrypted data drive, kept in a locked cupboard, within a locked office at the University of Central Lancashire. No copies of this Excel file were transferred to any computer used by the principle investigator. All other personal data (age, gender, domain identity, years of experience in graphic design, artwork ratings) were anonymised on an IBM SPSS file. This file was stored on a University of Central

Lancashire password protect staff laptop; a copy of this file was also backed-up on the UCLan staff drive area of the principle investigator.

6.3.2.1 Risk of Harm to Participants

One of the core pedagogies of graphic design education is the use of a critical review (commonly known as a "Crit"). During a crit, students present their finished artworks and have these commented upon by staff and students. From week one, of year one, student's at UCLan graphic design take part in a Friday Crit on previous work (such an assessment culture is not uncommon for other UK graphic design courses, or design education more broadly). Frequently, such crits make comparisons to other works under review, and relative opinions of merits and weaknesses are expressed.

It is the work under review rather than the designer of the work, but the results of this study using student artwork could reinforce self-labelling and issues of social standing within a peer group already established via crits. This was mitigated by a number of measures:

1. The result of the study would not use illustrations of any artwork provided. Thus, it would not be possible for the students taking part to know how judges rated their artwork from the results of the study.
2. The student artwork collected received an anonymous code at the start of the research. Only the principle investigator knew which codes related to which students, and the principle investigator has no involvement with graphic design students' assessment during their course of study. This information was stored in line with the DPA guidelines (see above) and was fully anonymised once the artwork has been stratified.

6.3.2.2 Consent

Consent was sought via participants signing the forms in Appendix A. Any participant who wished to withdraw from the study was informed they could contact the principal investigator to this effect. In the case of graphic design student's artwork, they were made aware they could withdraw up to the point of judges assessing the artwork. In the case of professional graphic designers judging the artwork, they were made aware they could withdraw up to the point of the data analysis.

6.3.3 Tasks

The Watching Words Move Task from study three was used and is called the Type Task from this point onwards. A new Image Task was used that required participants to select two images, and when seen side by side made some sort of creative juxtaposition, or interesting visual communication about the images chosen. As with the Type Task, this has a pedigree within graphic design education, and the juxtaposition of images is also a technique frequently used by professional graphic designers.

Whilst none of the artworks gathered for this study are presented here (as justified above), examples of six artworks previous and post this study are shown below in Figure 19, to give a visual context for these tasks. Specifically, the first Type Task example took the word "Coffee" and through the choice of typeface has rotated the letter "C" to resemble a coffee pot; the second Type Task example, based on the word "Imagine" has deleted the middle letter, leaving the viewer to imagine what this deletion may be; the third type example, has taken the word "Saw", and through extending the last letter "W" implies the teeth of a saw; the first Image Task example placed a pipe wrench next to an image of Robocop; the second a spiral shell next to a satellite image of a tornado; the third, the details of a leaf next to a section of road map.



Figure 19: Six examples of artwork (none of these examples were part of this study).

6.3.4 Instruction to Judges

For each task, whether type or image, two sets of instructions were developed (Table 25).

One set was an adapted version of Kaufman, Baer, Cole & Sexton's 2008 study, and is cited as an exemplar of CAT instructions (Kaufman, Plucker & Baer, 2008)

Table 25: Instructions for judges

--

Please look through these artworks, and rate them for creativity. There is no need to explain or defend your ratings in any way; we ask only that you use your own sense of which is more or less creative (relative to the other artworks provided).

Please look through these artworks three times, and rate them for creativity.

The first time familiarise yourself with all the artworks provided.

The second time, group the artworks into Low, Medium, or High ratings.

The third time, assign a numerical rating between 1 and 6 (1's being the least creative and 6's being the most creative).

There should be a roughly even number of artworks at each of the six levels. It is very important that you use the full 1-6 scale.

The other set of instructions were exactly the same as Kaufman's et al., with one difference, the first sentence "Please look through these artworks, and rate them for creativity" was replaced with two sentences adapted from Baer's 1993 CAT instruction for judges (a study where judges were specifically requested not to consider other factors that may impact on storytelling, for example, aesthetic appeal, or grammar, as part of their criteria for rating creativity). Baer's (1993) study is also cited as an exemplar of CAT instruction (Kaufman, Plucker & Baer, 2008). The wording used was as follows:

There is only one criterion in rating these artworks: creativity. We realise that creativity probably overlaps other criteria one might consider (for example: aesthetic appeal, or technical execution) but we ask you to rate the artworks solely on the basis of their creativity.

In Table 26, the relationship between the adapted wording described above and the original wording, on page 64 and 65, by Kaufman et al. (2008) was parsed to highlight what additions, deletions and adaptations occurred within Study 4.

Table 26: Comparison of instructions with Kaufman et al.

Kaufman et al. 2008	Study 4 (main instructions)
Please look through these poems twice	Please look through these artworks three times, and rate them for creativity.
	The first time familiarise yourself with all the artworks provided.

The first time, assign a Low, Medium, or High rating.	The second time, group the artworks into Low, Medium, or High ratings.
The second time, assign a numerical rating between 1 to 6, with 1 being the least creative and 6 being the most creative.	The third time, assign a numerical rating between 1 and 6 (1's being the least creative and 6's being the most creative).
There should be a roughly even number of poems at each of the six levels, but the numbers needn't be exactly the same	There should be a roughly even number of artworks at each of the six levels.
It is very important that you use the full 1-6 scale, however, and not assign all poems the same rating	It is very important that you use the full 1-6 scale.
There is no need to explain or defend your ratings in any way; we ask only that you use your expert sense of which are more or less creative.	There is no need to explain or defend your ratings in any way; we ask only that you use your own sense of which is more or less creative (relative to the other artworks provided).

Table 27: Comparison of instructions with Baer

Baer 1993 in essential of creativity (2008). (original quote, p.65)	This Study (caveat inclusion)
There is only one criterion in rating these tests: creativity	There is only one criterion in rating these artworks: creativity
I realize that creativity does not exist in a vacuum, and to some extent creativity probably overlaps other criteria one might apply	We realise that creativity probably overlaps other criteria one might consider
– aesthetic appeal, organization, richness of imagery, sophistication of expression, novelty of word choice, appropriateness of word choice, and possibly even correctness of grammar, for example-	(for example: aesthetic appeal, or technical execution)
but I ask you to rank the stories solely on the basis of your thoughtful-but-subjective opinions on their creativity	but we ask you to rate the artworks solely on the basis of their creativity.
The point is, you are the expert, and you needn't defend your choices or articulate a definition of creativity	
What creativity means to you can remain a mystery – what I want you to do is to use that mysterious expert sense to rank order the stories for creativity	

The rationale for the use of Kaufman's et al. instruction was to provide a clear procedure for judges to follow and to use instructions that were rating focused rather than ranking

focused, as Baer's instructions cited above are (though this does change in subsequent studies related to his 1993 work).

Subsequent additions to Kaufman's et al. wording were emphasised to clarify the issue of standards; a topic that arose during a post-assessment debriefing with graphic designers (and a topic present in the Study 3). Relative to other artworks, the instruction to rate the "most creative" seemed to be difficult for some judges. Setting aside issues of whether, in Study 3, the sample of artwork reflected enough creative diversity, the likelihood for such difficulty was the internal standards a judge had for creativity, and the congruency or discomfort experienced when they were tasked to label artworks "high" or "most" creative when, relative to the general standards of the domain, they were weak. Clearly, CAT instructions ask for comparison within a group of artwork (as does the theoretical assumptions which underpin this assessment technique) but it could still be the case that for graphic design professionals, mindful of a domain that judges them and others against expectations of high creativity, this is difficult to "turn off". In this respect, it was felt important to emphasise the relative nature of the ratings, and thus the phrase "relative to the other artworks provided" was added to supplement Kaufman's et al. instructions. Additionally, as the CAT is a relative rating, it was considered useful for judges to familiarise themselves with the range of artworks before beginning their assessment. This inclusion was a practical reinforcement that ratings were to be relative to the other artworks. Thus, the first stage of the judge's instruction for Study 4 asked judges to "...familiarise yourself with all the artworks provided" prior to beginning any ratings.

In contrast to Kaufman's et al. instructions, the key distinction to highlight for Baer's instructions is the emphasis his wording placed on one criterion: creativity; and explicit instructions that other factors related to creativity, such as technical execution and

aesthetic appeal, were to be discounted. This is a feature absent from Kaufman's et al. instructions. Moreover, with Baer's wording, he instructed judges to use their own "mysterious expert sense" (Kaufman, Plucker & Baer, 2008, p.65) of what creativity is, but to isolate this from other related factors (which he lists). In this way, Baer maintains only one unambiguous criterion is present.

This instruction could, however, be viewed as something of a contradiction; does such an instruction not imply other criteria are present and at work: to the extent that they need to be separated from creativity? Like general standards, as mentioned above, the ability to "turn off", to isolate creativity from other criteria, may be more challenging for some judges and straightforward for others. This potential for difficulty, however, need not infer that judges are unable to undertake such assessment: what Baer's original instructions acknowledge is the possibility that this may be a challenge, but one that is necessary and achievable. In this study, Baer's instructions were edited to two sentences and reflected a more formal tone in keeping with Kaufman's et al.

It is worth noting that whilst Kaufman's et al. instructions do not discuss other criteria (such as technical execution or aesthetic appeal) and Baer's make a considerable point of this, previous researchers using CAT have instructed judges to rate creativity alongside, but separate to, aesthetic appeal and technical execution (Hennessey, 1994). Indeed, in Amabile's 1982 work the extent to which creativity may be isolated from such factors was a formative part of her paper, and she concluded that "...although judges were not provided with a definition of creativity...they consistently and reliably identified a quality in both types of product that was distinct from technical execution. Moreover, for artworks, it was distinct from aesthetic appeal as well" (p.1010). At the same time, Amabile acknowledged that for some domains the distinction between technical execution and

aesthetic appeal may be less clear and that creativity is likely to correlate with these aspects of the work. Even within Amabile's 1982 studies, she found correlations as high as .77 between creativity and technical goodness, and Hennessey (1994) presented statistically significant correlations as high as .71. Contemporary CAT based research and studies more directly related to professional design, such as Valgeirsdottir, Onarheim & Gabrielsen (2015), have also identified high positive correlations, yet other design creativity researchers do not appear to have considered correlations in their research design (Lee, Gu & Ostwald, 2015).

A suggestion, for some time, has been that when a CAT is developed for a different domain, researchers should ask judges to rate both technical execution and aesthetic appeal (Amabile, 1982), and check to see creativity ratings are distinct from these criteria. Once this has been shown to be so, researchers need only ask for ratings of creativity, and can assume that technical execution is no longer a consideration for this task. It is at this point, as stated above, that Baer's instructions to discount technical execution contrast markedly with those of Kaufman's et al. The question is how much does that contrast matter?

6.3.4.1 Graphical Reinforcement of the Instructions and Rating Process

The rating sheet from Study 3 was redesigned, the purpose of which was to reinforce the textual instruction given, but offer more flexibility in the exact number of items placed at each level. Specifically, the adapted Kaufman et al. instructions stated "There should be a roughly even number of artworks at each of the six levels. It is very important that you use the full 1-6 scale." Based on text alone, as was highlighted in Study 3, not all judges follow this instruction, and a rating sheet was created that gave little scope for a judge to not follow this instruction. On reflection, this extreme was thought to be problematic, and a rating sheet was developed that aimed to offer a midpoint between giving no visual cues

to one that was too rigid, the aim was to visually help judges follow the CAT protocol, by clarifying the sorting and clustering of artworks to levels.

Three concepts were designed, see Figures 20 to 22, with the intention that these would be scaled to fit on an A3 sheet of paper, and be used in conjunction with 30 small-rectangular thumbnail images of the artwork being assessed. Thumbnails would be moved around the rating sheet by a judge, using the designated areas to sort and rate artworks.

For each option circles were chosen, rather than squares (as had been used in Study 3), as circles were considered less prescriptive about the number of artworks for each level. Option 1, was a concept to allow sorting to take place in three areas, and then subdivision into 1 to 6. In practice, the subdivision did not allow enough space for the thumbnails. In terms of the protocol, it also did not aid clarification of first sorting into low, medium and high, followed by rating from 1-6. For this reason, Option 2 and 3 were developed, and option 3 was used as it appeared to give the most amount of space for thumbnails, and was more visually simple.

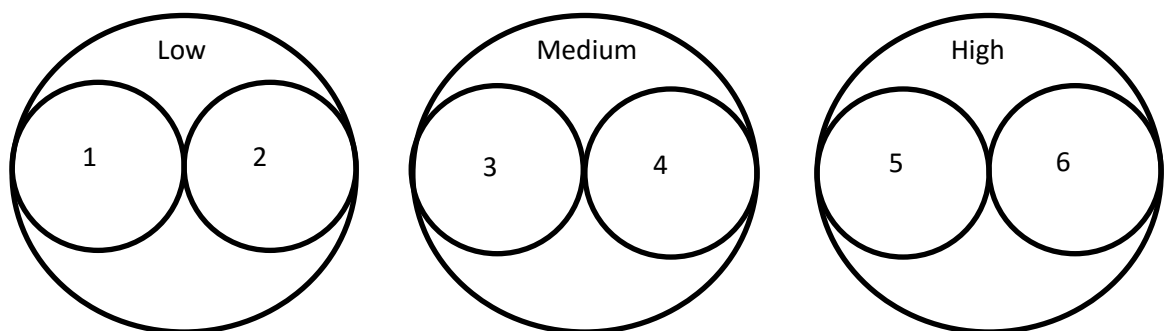


Figure 20: Option 1

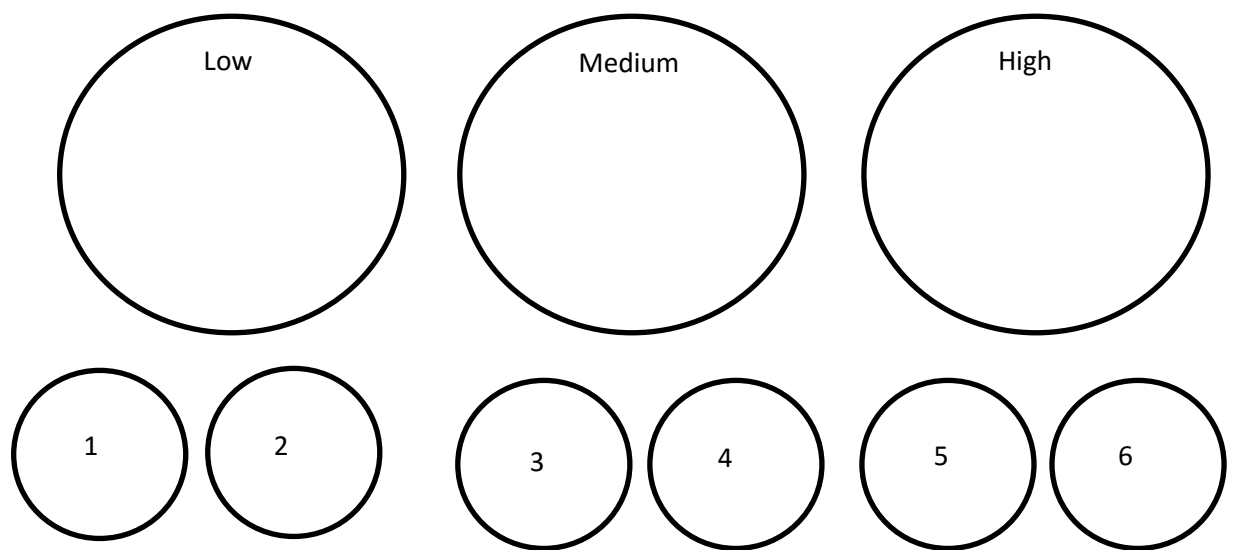


Figure 21: Option 2

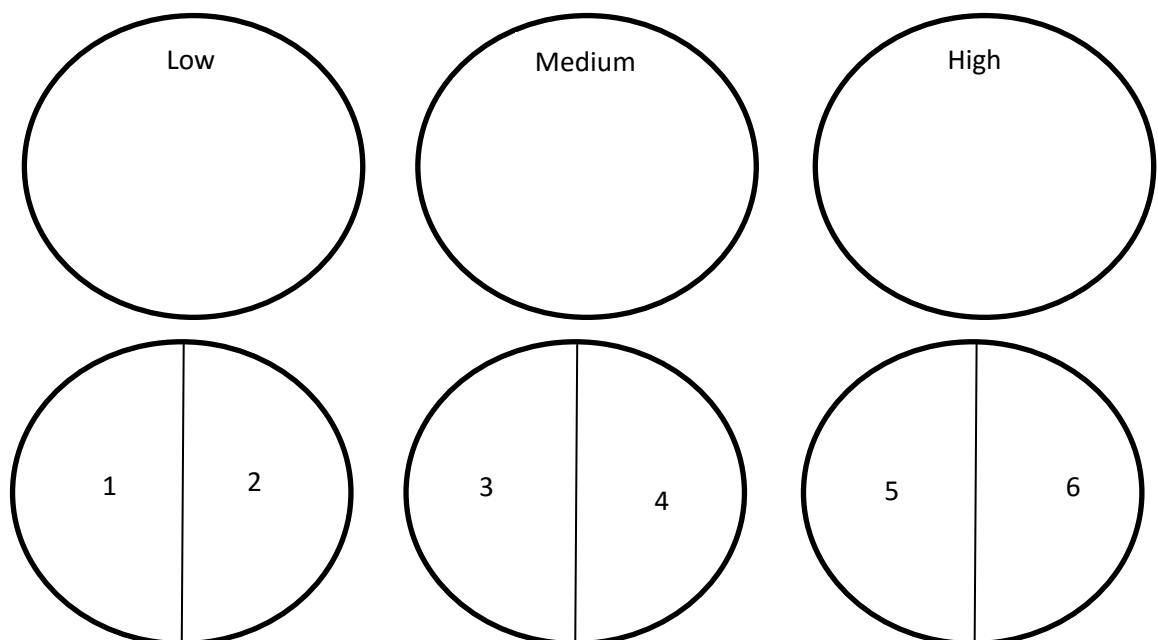


Figure 22: Option 3

6.3.4.2 Randomised Presentation of Artwork and Other Details

The rating sheet was used in conjunction with a full-colour high-quality display of the artwork, and a pack of relatively low-quality colour thumbnails of each artwork (that could be moved around on the rating sheet).

A number of options were available for full-colour display. A key consideration was the randomization of artworks shown to each judge, and loose leaf A4 prints seemed a practical route to take (allowing a researcher to reorder the randomised sequence prior to seeing each judge). Unfortunately, despite the high quality of the photographed artwork, the image quality was significantly less than when the artworks were shown as digital images in PowerPoint. The challenge of moving to PDF created, however, the need for individual PowerPoints to be manually re-ordered so the artwork followed the randomised sequence for that judge (see randomisation detail below).

A solution was found by using the Print to PDF option (PDF was also a better format as it did not need f5 to start it, nor did it end itself after the last image was shown). The crucial value was in using the randomised strings for each judge as the page numbers order to be printed as the PDF document. The creation of the strings, in practice, was slightly cumbersome, but not as great as manually re-ordering the PowerPoint slides. (To create the strings re-tabulate the sets to contain 1 to 30, using the paste special option in excel. Then save this file as a text file. Open this in another excel sheet, and deselect delimitators, and choose text. Then copy this string to Word, and replace spaces with commas to create the string to add to the PowerPoint print to pdf option.)

A further consideration was how to crop photographed artwork to suit the screen size of the PowerPoint file. A standard breadth was selected (297mm), which meant that some artworks from the Type Task did have enough height. As Photoshop files, the solution was

to increase the background to be 210 mm. Unfortunately, the increase in the white background often made the off-white of the cropped image stand out: readjust for brightness without undermining the quality of the typography was required in most cases.

6.3.4.3 Creation of Thumbnails

Where individual PDF files were created for each judge, this was not required for the randomization of thumbnail images. A set of artworks were organised onto an A4 sheet using Adobe Illustrator. This sheet was laminated (using matt sheets), and after careful attention to cutting out, only one set of thumbnails were required for each task, for all participants. Furthermore, lamination appeared to aid moving the thumbnails around the rating sheet. Codes were added on the back of each thumbnail, but these were not needed. The initial idea was that a number on the back would aid sorting into the randomised sequences and for recording a judges rating on the rating sheet. In practice, thumbnails were clear enough to identify and match to the PDF files for each judge, and a photograph of a judges rating sheet and the order of thumbnails was taken at the end of each task.

6.3.4.4 Selection of Artworks for this Study

Within a year group of 66 students, 48 students gave their consent to take part in the project. The median age was 19 years ($SD = 1.46$); 18 female and 30 male students took part. In order to select 30 artworks for each task, academic grades for each artwork were used. Grades were created as a result of student artwork for a creative thinking module: the criterion was creativity, and this was assessed by academic staff independent of this study.

The purpose of the selection process was threefold: firstly, to identify a diverse range of artworks across all CAT levels 1 to 6; secondly, to have five artworks represented at each CAT level; thirdly, to have the same participant represented in both tasks: type, and

image. The intention was that artworks would be paired so that the same person who completed the Type Task was represented in the Image Task (and vice versa). This would allow CAT scores for type and image to be aggregated to get a total score for both tasks.

For the purpose of artwork selection, there were two sets of academic grades available for each artwork. Firstly, the formative grade received after a critical review, and, secondly, the summative grade received for the end of semester portfolio. It is not uncommon after formative feedback for students to amend and improve their artwork for the end of semester portfolio. Indeed, this is encouraged by academic staff, and thus, in practice, it was possible to have four pieces of artwork and four grades available for each participant: a type artwork and image artwork for the critical review, and a type artwork and image artwork for the end of semester portfolio.

Grades from the critical review were ordinal rated (D- to A+) in contrast to portfolio grades, which used a ratio scale (0-100). To be able to compare these grades, the ordinal ratings were converted to an estimated scale grade as follows: minus grades were considered to fall within the lower 1/3 of a banding (i.e. 42, 52, 62, 72). A, B, C or D reflected the middle banding (i.e. 45, 55, 65, 75), and any plus grades would fall within the top 1/3 banding (i.e. 48, 58, 68, 78).

The CAT ratings for this study would require judges to rate artwork as low, medium, or high, and then rate these from 1 to 6. To determine which artworks and participants would be selected, each academic grade was stratified to a CAT rating as follows: marks between 44 and below, and 44-50 were rated as low (CAT level 1, and 2); 50-54, and 55-59 as medium (CAT level 3, and 4); 60-69, and 70 and above as high (CAT level 5, and 6).

As a result, representation at each of the CAT levels 1 to 6 was not equal for each task; some levels had much more artworks available than others. Because participants could have two artworks for each task, and only one was required, the data was cleaned for repetition. This was an iterative process, where priority was given to the least represented CAT levels first, the next least, and so on (for the type artwork this priority was CAT level 1,6,2,4 3, 5, and for image artwork this priority was CAT level 6,5,4,1,3,2). The result was that each task had one artwork per participant (48 artworks for type, and 48 artworks for image).

It is useful to highlight, that the purpose of this selection process was threefold: firstly, to identify a diverse range of artworks across all CAT levels 1 to 6; secondly, to have five artworks represented at each CAT level; thirdly, to have the same participant represented in both tasks: type, and image.

When each of the 48 artwork options were placed alongside each other, this highlighted some CAT levels had fewer options than others. For example, only five artworks were available at CAT level 1 for the Type Task, and only five artworks were available at level 6 for the Image Task. The inclusion of a participant in the Type Task at CAT level 1, for example, determined their representation within the Image Task, and vice versa. Moreover, when a participant was chosen to represent a specific CAT level (regardless of which task), this influenced the options available for other CAT levels. In this respect, the choices for selection became to an extent self-identifying, with limited options depending on whether a CAT level had more than five artworks available. The process was iterative and became progressively more challenging with each inclusion. However, the stratification was achieved, and 30 participants were identified whose text and image

artworks represented each of the six CAT levels, with five artworks at each level. For reasons of research ethics and participant confidentiality, examples of the artwork used in this study have been withheld.

6.3.5 Judges

Previous research guidelines suggest that for CAT reliability, between 5 and 10 judges is an acceptable number for a given task. For this study, 16 judges were required: 8 for the adapted Kaufman et al. instructions, and 8 for the adapted Baer instructions. A list of 49 graphic designers, within a 10-mile radius of the university campus was developed; emails were sent to each candidate and followed with a phone call two to three days afterwards. Judges were given the option to undertake the research at the university, or at a location of their choice (most often their design studio). 16 professional designers consented to take part, and undertook the CAT assessment of artwork as described below; six were female, and 10 were male. The mean age was just over 41 years of age ($SD = 9.80$) and ranged from 30 to 63 years of age. All judges identified themselves as graphic designers, and four judges suggested the following terms: Art Director, Creative, Creative Director, Graphic/web designer, tech. Each judge had over two years of professional experience in graphic design, the mean being just over 17 years, and ranged from 7 to 35 years. The majority of judges began professional work in graphic design at 22 years of age, but one judge began professional graphic design at 16 and another at 47.

6.3.5.1 Counterbalanced Conditions

Judges were randomly assigned to matched-pairs (Table 28) in which task and instructions were rotated to counterbalance whether they receive the Image Task first or second, with whether they received Kaufman only instructions (K), Baer only instructions (B), or a combination. This was to minimise order effects, particularly practice effects and fatigue effects.

Table 28: Counterbalanced tasks and instructions

Judge	1 st Task	2 nd Task
12 & 4	Image/K	Type/K
13 & 16	Type/K	Image/K
7 & 2	Image/B	Type/B
9 & 5	Type/B	Image/B
3 & 10	Type/K	Image/B
15 & 6	Image/B	Type/K
11 & 1	Type/B	Image/K
8 & 14	Image/K	Type/B

It is important to state that whilst order effect is not a variable in this study, it has been counterbalanced within the research design. For example, of the eight judges that rated the Type Task with Kaufman instructions, eight judges rated the Type Task with Baer instructions. Furthermore, where two judges began with the Type Task and Kaufman instructions, two began the Type Task with Baer instructions. This counterbalancing was exactly the same for those judges that rated the Image Task artwork. Whether Kaufman instructions or Baer instructions, within and between the sixteen judges each instruction experienced the same rotations of task and order of task.

6.3.5.2 Randomisation of Artwork to Which Judges

Randomization of Type artwork: 30 Type artworks were coded from 10-39 and randomised into 16 sets (coded from Set A to Set P). Each set was then randomly allocated to each judge (see Appendix B).

Randomization of Image artwork: 30 Image artworks were coded from 41-70 and randomised into 16 sets (coded from Set A to Set P). Each set was then randomly allocated to each judge (see Appendix C).

6.3.5.3 Randomisation of Judges to Which Experimental Condition

Judges were randomised to specific groups (Table 29). Excel randomization feature (1 to 10,000) generated a number next to each judge. These random numbers were then put in ascending order. Judges coded from 1 to 16.

Table 29: Randomisation of judges into groups

Group	Judge	1st Task	Order of Artwork	2 nd Task	Order of Artwork
1	12	Image/S	Set F	Type/S	Set K
1	4	Image/S	Set G	Type/S	Set H
2	13	Type/S	Set F	Image/S	Set N
2	16	Type/S	Set A	Image/S	Set M
3	7	Image/N	Set L	Type/N	Set E
3	2	Image/N	Set A	Type/N	Set P
4	9	Type/N	Set B	Image/N	Set D
4	5	Type/N	Set I	Image/N	Set C
5	3	Type/S	Set N	Image/N	Set K
5	10	Type/S	Set M	Image/N	Set O
6	15	Image/N	Set I	Type/S	Set D
6	6	Image/N	Set E	Type/S	Set G
7	11	Type/N	Set C	Image/S	Set B
7	1	Type/N	Set J	Image/S	Set P
8	8	Image/S	Set J	Type/N	Set L
8	14	Image/S	Set H	Type/N	Set O

6.3.5.4 Data Analysis of Inter-Rater Reliability

The most commonly used measure of inter-rater reliability for CAT studies is Cronbach's alpha and, as with the previous studies, was the measure used in this study. To compute the alpha's for each task in relation to instruction, judge's ratings were collated as detailed in Table 30.

Table 30: Collation of judges rating for alpha calculations

Image Task/Kaufman	Judge 12, 4, 13, 16, 8, 14, 11 & 1
Image Task/Baer	Judge 7, 2, 9, 5, 15, 6, 3 & 10
Type Task/Kaufman	Judge 12, 4, 13, 16, 15, 6, 3 & 10
Type Task/Baer	Judge 7, 2, 9, 5, 8, 14, 11 & 1

6.3.5.5 Differences in Ranking/Score

Theoretically, it is possible that both groups of judges could have perfect alphas, yet have the perfect opposite rankings; i.e. the judges who received Kaufman's et al. instructions

agreed perfectly within their group on the rating for each artwork, but their ratings were the complete opposite of the judges who received Baer's instructions (who also had perfect agreement within their group).

Pearson's r is an option to consider the correlation between each group, and the strength of that relationship. If the assumptions for Pearson's r cannot be met, then Spearman's rho or Kendal's tau are non-parametric alternatives. In addition, scatter plots, and simple linear regression would provide data on which to consider score/ranking differences between groups.

6.3.6 Procedures

The procedures for rating artwork were the same for each judge and task. Initially, judges answered three questions: their years of experience in graphic design, whether they would describe themselves as a graphic designer and their age. After this, each judge was given the instructions for their first task, alongside an example of the design brief given to participants for this task. They had as much time as they required to read the instruction. Next, each judge accessed a laptop with a PDF slide presentation of the 30 artwork for the first task. The order of artwork was randomised, and they were free to control how long they viewed artworks, and could return to each artwork for further inspection. Each judge familiarised themselves with all the artworks, and when satisfied informed the researcher they were ready to continue. Judges were given an A3 laminated rating sheet, see Figure 23 (developed to reinforce Kaufman's et al. CAT protocol and instructions graphically), and a set of laminated cards.

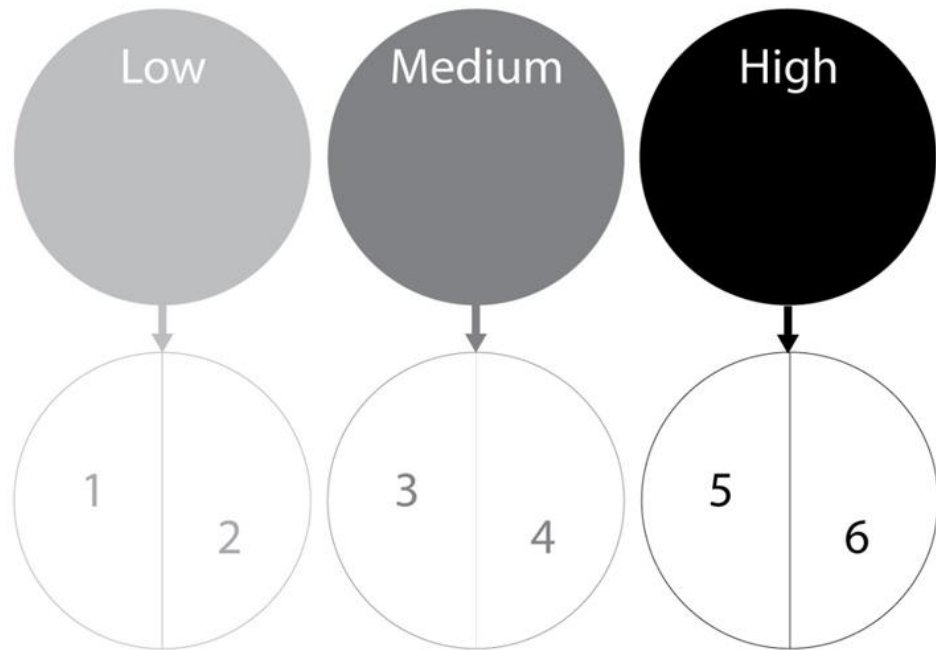


Figure 23: Rating sheet developed to graphically reinforce Kaufman's et al. CAT protocol

These cards were miniature copies of the artwork they had just viewed; the order of the cards followed the same order as the PDF slide presentation. Cards were placed in a stack, by the researcher, onto the rating sheet area designated "medium". Judges proceeded to rate the artwork and were given as much time as they required. Upon completion, the research took a photograph of the task 1 ratings. Task two followed the same procedures: reading instructions, viewing all 30 artworks for task 2 on the laptop, moving the thumbnails to reflect an assessment, and photography of the completed rating sheet. All judges were debriefed on the purpose of the experiment, and given the chance to ask any questions about the study.

6.4 Results

6.4.1 Type Task

For the Type Task, the 8 judges who received the adapted Kaufman et al. CAT

instructions had an alpha of 0.73; for the 8 judges that received the adapted Baer's instruction to discount technical execution the alpha was 0.82. The skew for aggregated ratings for each group was within acceptable limits at the less than 5% level: adapted Kaufman et al. z-skew was -0.62; adapted Baer z-skew was -0.2. Pearson's r was 0.82, suggestive of a very strong positive correlation between the scores, and was a significant correlation at the 0.01 level (two-tailed). The scatter plot below (Figure 24) shows the regression line and strength of correlation.

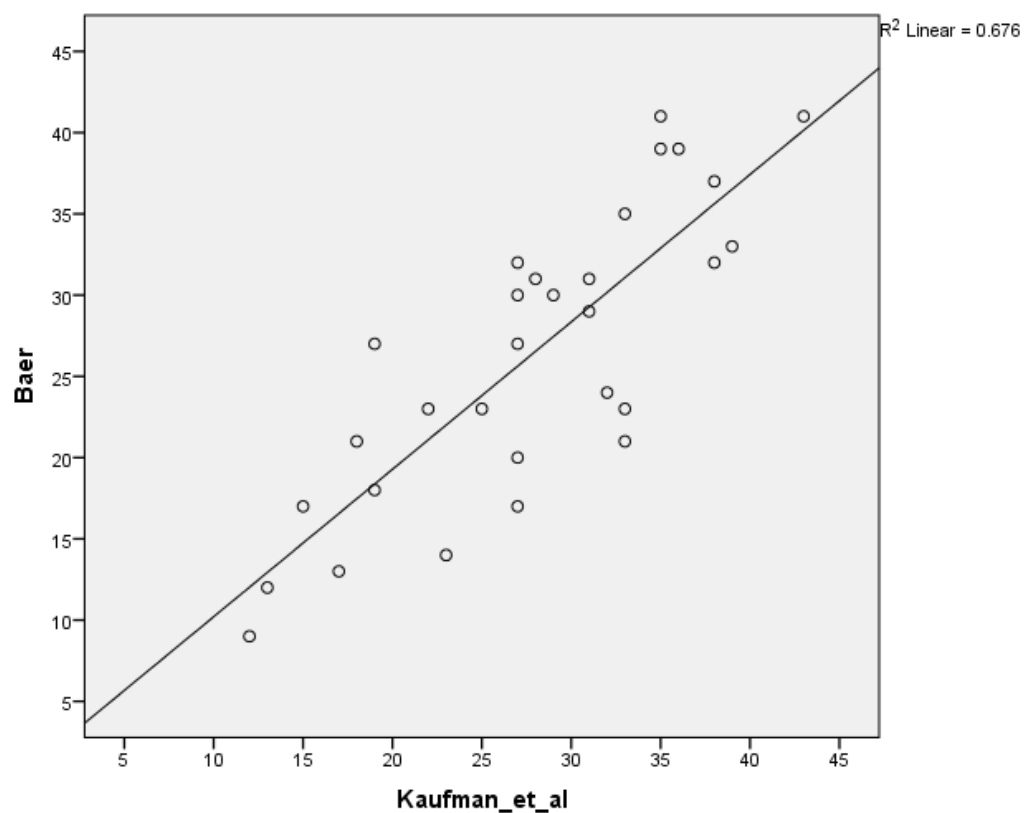


Figure 24: Scatterplot for Type Task

6.4.2 Image Task

For the task, the 8 judges who received the adapted Kaufman's et al. CAT instructions had an alpha of 0.75; for the 8 judges that received the adapted Baer's instruction to discount technical execution the alpha was 0.80. The skew for aggregated ratings for

each group was within acceptable limits at the less than 5% level: adapted Kaufman et al. z-skew was 0.74; adapted Baer z-skew was 0.10. Pearson's r was 0.62, suggestive of a strong positive correlation between the scores, and was a significant correlation at the 0.01 level (two-tailed). The scatter plot below (Figure 25) shows the regression line and strength of correlation.

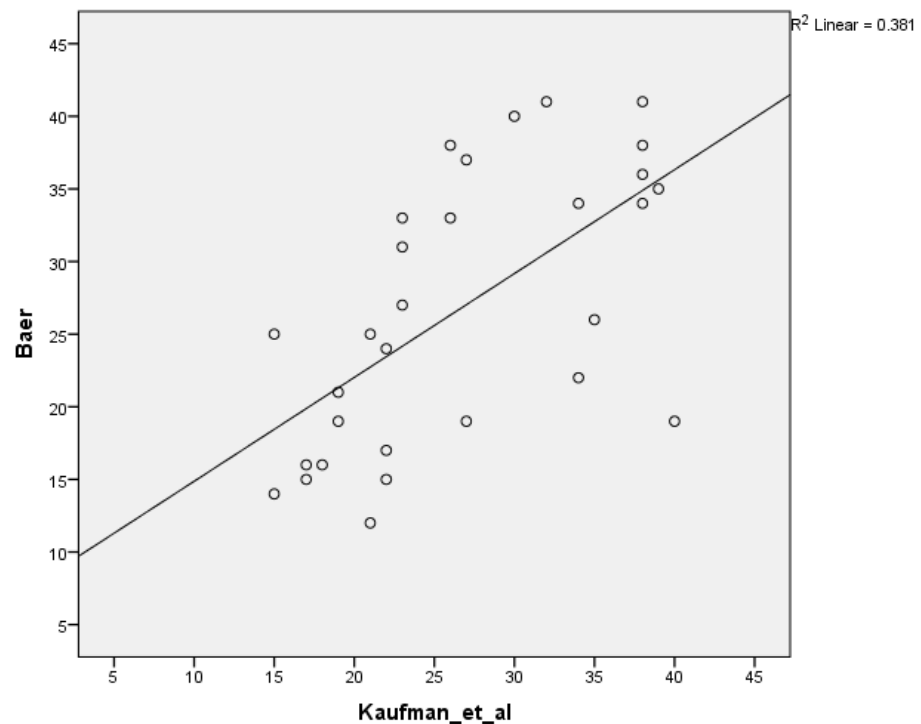


Figure 25: Scatterplot for Image Task

The correlation between the adapted Baer's instructions for both type and image was an r of 0.43, which can be considered a moderate positive correlation and one that was significant at the 0.5 level (two-tailed). For the adapted Kaufman et al. instructions for both type and image r was 0.11, suggestive of a negligible positive correlation, and one that was not statistically significant. Figure 26 highlights alpha and r in relation to which task and instructions.

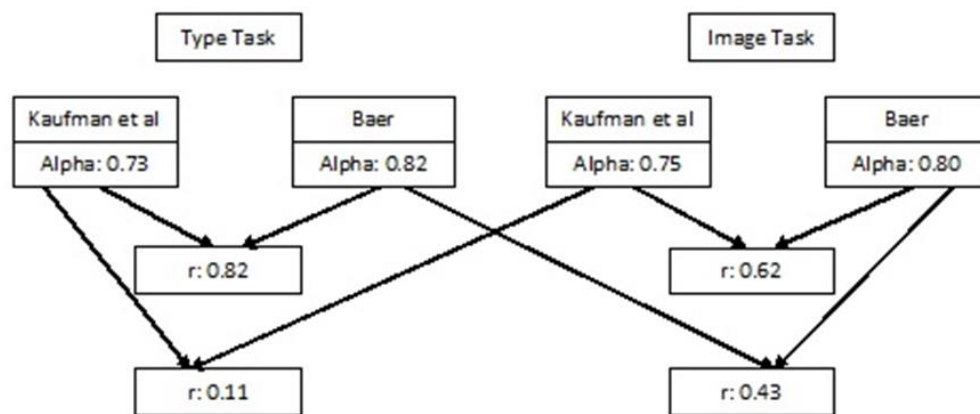


Figure 26: Alpha and r in relation to task and instructions

6.5 Discussion

It was expected that using the CAT to assess graphic design creativity would follow a similar pattern of inter-rater reliability as it had in other design domains. However, the results of Study 3 did not demonstrate this. A number of reasons, such as the range of artwork, technical emphasis for some judges, task selection, and the sampling of judges were considered, and accounted for in Study 4, and appear to have resolved the previous issue of inter-rater reliability. In either task, the level of inter-rater reliability was acceptable, all measures were above 0.7, the highest being 0.82. Given this, these particular tasks can be considered reliable measures of graphic design creativity using the CAT. Prior to this study, the choice of tasks to measure graphic design creativity was not obvious, and the Study 3 findings highlighted that some tasks do not translate well from design education context into experimental research.

Whilst each task (regardless of the instruction to judges) had acceptable alphas, a marked difference can be seen in the correlation between the Type Task ($r=0.82$) and the Image Task ($r=0.62$). The reasons for this difference could be numerous, complex and interrelated. Perhaps the Type Task, and typography, share a common knowledge base

for graphic designers; possibly the Type Task is less influenced by discounting technical execution than the Image Task; the inclusion of colour in the Image Task may add to the complexity of assessing its creativity: colour was absent in the Type Task. What can be said, and has been said by other researchers (Reiter-Palmon et al., 2009), is that task selection is an important factor in creativity assessment, and our depth of understanding is "essentially missing in the literature" (Lubart & Guigard, 2004, p.48). With Study 4 (and the broader discussion and findings of Study 5 to follow) the findings of this research reinforce this position in relation to graphic design creativity assessment. While these tasks are well-established measures of graphic design creativity in an educational context, their application to design creativity research would appear unpublished to date. But, in translating these tasks from an educational context to a research one it has resulted in challenges and insights to CAT protocol that have practical implications for design creativity assessment, and possibly creativity assessment as a whole.

6.5.1 Diversity of Creativity Within a CAT Grouping

Aside from the task itself, a further impact on levels of consensus could be the range of work being presented to judges. The case was argued above that because the CAT does not allow judges to confer on their assessment, it is possible it is more susceptible to skewed samples than appears to have been suggested in the CAT literature to date. Certainly, this argument can be made for highly creative artworks, and then by extension, the case can be made for other skewed samples be they predominantly low or medium in quality

Highlighting the lack of opportunity for discussion among judges is not to be taken as a weakness of research method, indeed, it has a number of benefits, but it does imply a potential limitation for CAT studies, or at the least something to reflect upon when

consensus is lower than expected (which was the situation in Study 3).

With this study, a significant effort was made to select works that reflected a range of creativity. The conclusion, given the findings of Study 3, and the subsequent study here, is that diversity of artwork may play a significant factor in levels of consensus amongst judges. Therefore, it is possible for researchers to assume enough diversity exists in a sample of artwork when it may not.

It is useful to emphasise that for some researchers familiar with CAT protocol, the selection of artwork for this study will be seen as problematic: artificial and unrealistic. The first point to acknowledge is they are absolutely correct. The selection of artwork is "artificial" and "not realistic" on purpose due to the experimental nature of this study. Understandably, a different synonym would be to state that the study was designed to "optimise" the CAT, as in the sampling was designed to be as close to ideal/optimal conditions as possible in order to evaluate the level of consensus amongst judges.

The rationale for this was partly due to the CAT being a relative assessment of creativity (relative to the other work in the sample), but mostly due to the fact that previous CAT studies in graphic design (both the findings of Study 3, and others studies like Wojtczuk, 2014) had not achieved a suitable level of inter-rater agreement. Given some consistency that CAT did not work in graphic design, why might this be the case was a key focus of Study 4; one specific area of consideration was the range of artwork. In this respect, it was considered prudent that some form of pre-test for diversity of artwork be used. This would offer a degree of control and insight into how diverse the artwork presented to judges would be (it became a controlled variable for the research design), and created optimal conditions under which CAT assessment would take place. As discussed

previously, if a suitable level of consensus could not be achieved under these optimal experimental conditions then, indeed, something deeply problematic was likely to be occurring in relation to the CAT and graphic design creativity.

With these optimal conditions, and the other changes discussed below, for the first time the CAT worked, in both cases on two well-regarded design tasks that aim to challenge graphic design creativity. The question remained, however, would the CAT work in less optimal conditions than those created in Study 4, that is, with artwork that had not been pre-selected? As will be discussed in the final study (Study 5) of this thesis, with the caveats included below, the answer was yes.

6.5.2 Technical Ratings

In early CAT research, Amabile (1982, 1996) concluded that judges were able to distinguish creativity from other aspects of the work such as aesthetic appeal and technical execution. Does this finding apply to graphic design creativity? Whilst the differences between the adapted Kaufman et al. instructions and the adapted Baer instructions (to discount technical execution) ranged from 0.05 to 0.09, the difference was towards higher levels of inter-rater reliability when judges were asked to discount technical execution from their creativity ratings; this occurred in both the Type Task and Image Task. Arguably, in the present study, these differences are slight, but it may be that acceptable inter-rater reliability is not enough in isolation and that the other consideration is the correlation between aggregated scores.

6.5.3 Rankings, not Just Alpha

In this respect, correlations between instructions and task show a greater level of difference. Whilst correlations were significant at the 0.01 level in either task; they were

stronger for the Type Task than the Image Task.

It appeared that the Type Task was less influenced by discounting technical execution, and, to an extent, the Image Task may be more susceptible. A further consideration was the correlation between type and Image Tasks relative to the instructions judges received. Only the adapted Baer instructions were statistically significant, and suggestive of a moderate to strong positive relationship (Fig. 29.).

6.5.4 To Include a Caveat or not?

The main point to consider is whether to include a caveat around technical execution in future research. By inclusion, such a caveat directly addresses assumptions around technical execution and creativity. Indeed, if the CAT is foremost a measure of creativity (however judges interpret this word), then a clarification on technical execution seems a reasonable distinction to bring to their attention.

The increase in inter-rater agreement for judges that received the technical execution caveat can be interpreted both for and against its inclusion. One argument is that these increases are slight (0.09 and 0.05 respectively) and could be due to sampling error, the other argument is that for two different tasks both alphas increased when a caveat was included, and given the novelty of the CAT as a measure of graphic design creativity, it is difficult to consider if such increases are within common fluctuations or could be on the margins of statistical significance.

More revealing are the correlations between type and image scores relative to instructions. It is only scores where a caveat was included that an aggregated graphic design creativity score was possible. As the purpose of each task was to evaluate graphic

design creativity by isolating two distinct features of graphic design (the creative use of type, and the creative use of image), the expectation was that a degree of positive correlation would be likely in the combination of these tasks. Indeed, this is much of the premise upon which module grades are established within graphic design education, particularly in the 1st years of a course: a portfolio of work is submitted at the end of a module, and it is the aggregated scores for each weekly task that determines the grade for that module. That said, this correlation need not be a strong one, it is quite possible for a person to show a high ranking on one task and less so on another.

What is interesting to note for Study 4 is that when judges assess exactly the same tasks and exactly the same artwork, the caveat on technical execution offers more assurance in an aggregated score, and thus, at least for research purposes, enables distinctions within a group on levels of graphic design creativity.

The last consideration and a salient point for clarification is why did this study not allow judges to assess creativity, and technical quality and aesthetic appeal. As a CAT protocol this is established in creativity research within experimental psychology (Amabile 1982, Hennessey, 1994) alongside design-based studies such as Christiaans & Venselaar (2005) and Valgeirsdottir, Onarheim & Gabrielsen (2015).

The rationale for not considering this form of CAT protocol was partly addressed in the method section (Study 4) regarding the instructions to judges but is worth directly addressing this again here.

For flow of discussion, the key points made previously were that whilst Kaufman's et al. instructions did not discuss other criteria (such as technical execution or aesthetic appeal)

and Baer's made a considerable point of this, previous researchers using CAT have instructed judges to rate creativity alongside, but separate to, aesthetic appeal and technical execution (Hennessey, 1994). Indeed a suggestion, for some time, has been that when a CAT is developed for a different domain, researchers should ask judges to rate both technical execution and aesthetic appeal (Amabile, 1982), and check to see creativity ratings are distinct from these criteria. Once this has been shown to be so, researchers need only ask for ratings of creativity, and can assume that technical execution is no longer a consideration for this task. It is at this point, that Baer's instructions to discount technical execution contrast markedly with those of Kaufman's et al.

Thus, what has occurred in many CAT studies is that a variety of possible instructions and protocols are open to researchers. This can be a valuable feature, and diversity and flexibility of method can partly explain the benefit of the CAT as a highly adaptable domain-specific measure of creativity. The less positive aspect is that core aspects of the method can change from one study to another, and from a scientific perspective that degree of interpretation is problematic. The position taken in this thesis is that it is also unnecessary; more is to be gained from the creativity research community having standard instructions for judges than is likely to be lost.

In this respect what this study explored was the standard instructions to be given to judges, regardless of whether a factor analysis had shown that technical execution or aesthetic appeal was distinct from creativity for a given task. The argument against Amabile, and other researchers who approach CAT assessment without providing Baer like instructions to judges, is that their instructions only imply for judges to discount technical execution and aesthetic appeal, rather than state this directly, as Baer's does

and were tested in this study.

This is a subtle point, but hopefully, a clear one to draw attention to. When a research study asks a judge to rate creativity, technical execution and aesthetic appeal, that judge is likely to conceptualise creativity as something different because the research study has asked them to give three different sets of rating for the same works. A follow-on study (depending on whether they use Kaufman or Baer's instructions) with different judges, may not make mention to technical execution and aesthetic appeal. As the research design has not prompted the judge to conceptualise a distinction, their instinctive concept of creativity may have more to do with technical execution. The answer to overcoming this seems a simple one, when the judge first assesses creativity, standard protocols should ask them not to consider technical execution as part of their rating of creativity.

For some researchers, this may sound like labouring a point, and possibly a minor one at that, but this perhaps only sounds obvious in hindsight. The context to remember is that, at the time of this study, no CAT research study in graphic design had achieved reliable alphas, and several had tried. What appeared to be the case was that a graphic design CAT was extremely sensitive to methodological choices like those stated above.

Clearly, precisely how researchers interpret the CAT method and protocol are open to debate (Amabile, 1982; Baer 1993; Hennessey, 1994; Kaufman, Baer, Cole & Sexton; 2008). In essence, the argument presented here is to develop standard CAT instructions to rate creativity for all design creativity researchers: those who choose to measure additional criteria, such as technical execution and aesthetic appeal, in each study; those who do so only once; those that measure only creativity from the start. In design domains where the distinction between creativity, technical execution and aesthetic appears to be a

rather than one, the findings of this study suggested that explicit CAT instructions to discount technical execution from creativity ratings are more reliable.

6.6 Conclusion

Prior to this study, the choice of tasks to measure graphic design creativity was not obvious, and the findings of Study 3 highlighted that some tasks do not translate well from design education context into experimental research. Therefore, there was little precedent established for a task to measure graphic design creativity using the CAT reliably. The issues as to why a CAT did not appear to work for graphic design raised both practical and theoretical implications for creativity research within and beyond graphic design

In this study, a number of research design factors, such as the range of artwork, technical emphasis, task selection, and the sampling of judges were accounted for in the research design and appear to have resolved the previous issue of inter-rater reliability. However, the arguments presented in this chapter suggest instruction to judges to discount technical execution from their creativity assessment does appear to influence the reliability of the CAT.

The difference was towards higher levels of inter-rater reliability when judges were asked to discount technical execution from their creativity ratings; this occurred in both the Type Task and Image Task used. Moreover, only CAT assessments undertaken where the technical caveat was included enable an aggregated graphic design creativity score for both the image and Type Tasks. Perhaps these implications apply not only to graphic design, but have relevance for all CAT assessments of design creativity?

To paraphrase Nickerson (1999), researchers have two choices: include a caveat on

technical execution that future research will show was not that important, or exclude it and find out technical execution does influence rating on graphic design creativity.

Unless there is some detrimental effect (which does not appear to be the case in this study), then a cautious approach would be for future CAT usage to include a caveat on technical execution when applied to design creativity research

Chapter 7, Study 5: The Relationship Between Graphic Design Competencies and Creative Output

7.1 Introduction to Study 5

...there is not a significant correlation between visual-spatial ability and graphic design ability. However, this finding is counterintuitive and may be due to the broad and complex nature of visual-spatial intelligence. The ability to recognize and understand shape, form, and space is essential to the visual arts, particularly graphic design. (Mackie, 2005, p.85).

Returning to the original intent of this research, based on findings in chapters 3 to 6, in Study 5 several focused questions about the relationship between skills, creativity and domain were explored.

With a reliable measure of creativity optimised for graphic design (Study 3 and 4), and a conceptual framework of competencies that integrated both applied and theoretical knowledge of graphic design abilities (Study 2) it was possible to test the relationship between specific competencies and creative output through empirical study.

In relation to graphic design creativity, Study 5 occurred within a live research context. Where Study 3 and 4 offered judges non-parallel artworks to explore CAT protocols, Study 5 acquired artworks created directly for this study. It also sought to explore further the influence of technical execution and aesthetic appeal on graphic design creativity, and the impact of researcher's verbally re-emphasising "creativity only" instructions when briefing judges prior to creative assessment.

7.2 Background to this Chapter

At the beginning of this thesis, it was highlighted previous research had identified linguistic intelligence as more important to graphic design creativity than visual-spatial intelligence (Mackie, 2005). The implications suggested that greater focus should be placed on the development of reading and writing abilities, and less emphasis on drawing ability. If correct, this inferred a significant re-evaluation of educational traditions within graphic design curriculum, and across secondary, further and higher education.

However, were these findings still valid when accounting for current research methods?

Very little empirical research had been conducted on graphic design creativity and specific cognitive abilities (of which Mackie's work was an exception), and since 2005, researchers understanding of how to measure cognitive abilities like mental visualisation has become more nuanced in relation to artists and designers. For example, as described in Study 2, the work of Kozhevnikov's et al. (2013) has shown that spatial visualisation tests reflect only part of our current understanding of visualisation in art and design (the tests Mackie used would, today, be considered biased towards more scientific and engineering backgrounds than design and artistic professions).

Additionally, in relation to assessing graphic design creativity, though this was not a direct focus of Mackie's research, he highlighted the challenges of portfolio assessment for his work, and within art and design research literature more broadly. This is understandable given the finding of this thesis through Study 3 and 4, and other researchers' experiences of applying the CAT to graphic design. Without optimal conditions, and the introduction of several caveats for CAT protocol, research had not established a reliable measure of graphic design creativity for a single task yet alone a portfolio. With optimal conditions, Study 4 suggested acceptable levels of consensus were possible for specific tasks, however, this form of graphic design creativity assessment had yet to be tested in less

optimal conditions.

7.2.1 A Range of Graphic Design Experience

One of the caveats of Study 4 was that diversity of artwork might play a significant factor in levels of consensus amongst judges, and for researchers not to assume enough diversity existed in a sample of artwork (especially when participants were drawn from a single population, i.e. a student year group). Where Study 4 made a significant effort to pre-select artworks that reflected a range of creativity within a student population and provide optimal conditions for judges, this was not the approach taken in Study 5. Here the research design was intended to be more closely related to that typically used for CAT assessment. For example, the recruitment of participants to undertake a creative task in addition to other tests.

As no pre-selection of artwork would take place, a plausible way to be confident that enough diversity of graphic design creativity would occur was to stratify the sample according to levels of graphic design expertise. For example, within the categories of novice, intermediate and expert, it was likely that a range of graphic design creativity fell within and across these categories. It was plausible that experts and intermediates were likely to produce more creative graphic designs than novices (and this was the logic for the stratification). While this may be a fair assumption, it was also one that may not occur in some instances. A novice, for example, may have a talent for graphic design creativity that could exceed that of a student of graphic design; an expert could have had a number of pressing deadlines at the time they created their artwork and be rated lower than a competition winning graphic design student relatively free of such professional commitments.

These types of situations were possible within such a sample, but with an appropriate size of group, such individual difference could be accounted for within the group as a whole. What was important was to offer a range of graphic design creativity to judges; it was the group as a whole that was the basis of a dependent variable of creativity. With whom creativity was associated, be they novice, intermediate or expert was not a concern. Indeed, for ethical reasons, and to aid with the recruitment of professional designers, it was not a consideration of Study 5 to identify or evidence the relationship between graphic design experience and creativity. The aim was to explore whether the CAT would remain reliable without the pre-selection of artwork and, if it did, to what extent graphic design creativity correlated with specific cognitive abilities.

7.2.2 Graphic Design Creativity and Novice Participants

By stratification of participants into novice, intermediate and expert, one challenge was whether the creativity tasks and specifically the Type Task, used in Study 4 could be successfully completed by participants who had no background or experience in graphic design. For this reason, prior to the recruitment of participants, both the Image Task and Type Task were piloted with novices.

As was identified in Study 3, the results from the T-shirt graphic design task highlighted that the inclusion of hand-drawn artwork alongside computer-aided artwork likely introduced an additional challenge for judges when evaluating creativity. Indeed, it was for this reason that the subsequent Study 4 only focused on artwork that had been created on a computer, and this, along with other caveats, appeared to have improved the level of inter-rater agreement.

While both experts and intermediates were likely familiar with either the type or Image

Task through their professional and educational experience, those who were not would certainly be versed in the professional computer software that can aid undertaking the Type Task.

For the most part, it is through computer software packages, such as Adobe Illustrator or InDesign, that much graphic design artwork is created. In relation to these specific packages, it was reasonable to assume that experts and intermediates will have few issues with gaining access easily. As this was not likely to be the case for novices, this raised the issue again regarding consideration for task design and implications of not gaining access to these industry standard packages for novices. A more pressing consideration was that such packages were not straightforward to use, and the learning curve would be a significant undertaking for a novice without some form of training. It was also questionable if that level of commitment towards a research study was a reasonable request for a researcher to make.

This is not to say the Type Task cannot be undertaken with less sophisticated software, but it is dependent on the concept a participant had in mind. At a certain point, what is relatively easy to sketch by hand is not so straightforward to implement in a basic software package (in such instances, inbuilt limitation regarding the visual manipulation of text mean such software has become the wrong tool for the job).

To not acknowledge the probable mismatch in technical software skills between novice and intermediate/experts would introduce a research design bias that could impact on either group. For example, if only basic computer software was used, whilst this may match the technical skill level of the novices, it could undermine the creative freedom of the intermediate/experts. Perhaps of more importance was that the limitations of basic

computer software might restrict the creativity for all participants, as the challenge would become not about one's creative vision, but how to implement this within the constraints of the software. For example, Microsoft Word and PowerPoint were software packages available to the large majority of working participants, yet the use of this software for graphic design would be unusual and frustrating for a professional graphic designer familiar with Adobe Illustrator. It is not simply that professionals would need to learn to use PowerPoint to replicate Adobe Illustrator, and in this way, one could argue, follow a similar learning curve as novices, but that their past knowledge and practice placed them at a disadvantage to novices. For example, Frensch & Sternberg (1989) found in a study of expertise, that expert Bridge players were disadvantaged by fundamental changes in the rules of the game compared to novices who were unfamiliar with the traditional rules and learning the game for the first time. While learning to use Microsoft Word for the Type Task was not the same context as Frensch & Sternberg's expert Bridge players, the precedent of disadvantaging one group of participants over another through task design could work against experts and intermediates as much as novices.

What was required for the Type Task was to reduce this mismatch in technical software skills; to the extent that not knowing how to do something in a computer package did not impinge on the creativity of the concept, and its resulting manifestation to the satisfaction of the participant.

In contrast, the Image Task, with the juxtaposition of two cropped images, is less influenced by the need for specialist illustrative software, and could as easily be achieved by basic Windows programs like Paint, or Microsoft Office applications like PowerPoint or Word. It would seem given the proliferation of these programs, that most participants would be familiar with the basic functions of these ubiquitous applications.

As two-thirds of the participants were intermediate or experts, the decision was taken not to restrict the software they would have available for the Type Task used in this study. To make sure that novices were not disadvantaged, the solution, then, was to give novices access to technical software skills that will enable them to realise their ideas. In this respect, novices may not know how to produce their ideas in Illustrator, but would be able to dictate their ideas to a technician who could. It may be that such software support could be valuable for intermediates and experts, but it is of foremost importance for novices.

Exactly how this support was to be provided, for how long, and how to guard against software support slipping into design suggestions, and thus not reflecting the creativity of the participant, were considered. For example, decisions on layout, type and choice of work would belong to the novice, with the technical role purely to transfer ideas into Adobe Illustrator. Such conditions were presented to pilot participants through the information on technical support (Figure 27 & 28).

Information on Technical Support for the Image Task

As this Image Task is designed to offer a creative challenge across the full range of graphic design expertise, it is important that the tools available to create artwork provide the greatest freedom of expression.

To this end, one industry standard software used by professional graphic designers is Adobe Illustrator, and this software will be available, alongside technical support for using the software for each participant in this study.

The technical support is to enable each participant, regardless of expertise with Adobe Illustrator, to create an artwork that reflects his or her creative vision. It is vital however, that this technical support only deals with the practical aspects of transferring a participants design to Adobe Illustrator, rather than offering advice or suggestions on the quality of the design.

To safeguard against such influence technical support will be offered remotely via email. For example, a participant may send two photographs/images via email to researcher's personal e-mail. These images will then be transferred to Adobe Illustrator, and this digital illustration will be sent back to the participant. Any amendments required

by a participant can be suggested, and the cycle will continue until the participant is satisfied, or until the time limit on the task has expired.

The time required to create this artwork is expected to take no more than one hour. However, thinking time and playing with ideas may increase for some participants.

This task is not timed in a conventional sense but for practical purposes a limit of one week from receiving the task has been set. How productive or efficient graphic design ideas can be created is not part of this study, and one week will allow participants the chance to think, reflect and play with their ideas prior to submitting a single artwork via email, or through the technical support option described above.

A time limit of one week will allow participants' ideas to incubate, which is often an opportunity lacking in many experimental creativity research studies. The reason for this is the desire to control for the influences of other individual's ideas on the participant's own work. To mitigate this influence, we ask that you do not discuss your ideas with others prior to completion within the week. In this way, artwork can only be attributed to your own efforts, which is exactly what is required for this study.

Unless otherwise stated by a participant, the size of the image of both images will be cropped to 10 cm by 10 cm, and will be cropped relative to the centre of the image.

Figure 27: Information on Technical Support for the Image Task

Information on Technical Support for the Type Task

As this Type Task is designed to offer a creative challenge across the full range of graphic design expertise, it is important that the tools available to create artwork provide the greatest freedom of expression.

To this end, one industry standard software used by professional graphic designers is Adobe Illustrator, and this software will be available, alongside technical support for using the software for each participant in this study.

The technical support is to enable each participant, regardless of expertise with Adobe Illustrator, to create an artwork that reflects his or her creative vision. It is vital however, that this technical support only deals with the practical aspects of transferring a participant's design to Adobe Illustrator, rather than offering advice or suggestions on the quality of the design.

To safeguard against such influence, where possible*, technical support will be offered remotely via email. For example, a participant may send a hand drawn illustration of their design via email to kjeffries@uclan.ac.uk. This hand drawn illustration will then be transferred to Adobe Illustrator, and this digital illustration will be sent back to the participant. Any amendments required by a participant can be suggested, and the cycle will continue until the participant is satisfied, or until the time limit on the task has expired.

The time required to create this artwork is expected to take no more than one hour. However, thinking time and playing with ideas may increase for some participants.

This task is not timed in a conventional sense but for practical purposes a limit of one week from receiving the task has been set. How productive or efficient graphic design ideas can be created is not part of this study, and one week will allow participants the chance to think, reflect and play with their ideas prior to submitting a single artwork via email, or through the technical support option described above.

This time limit of one week will allow participants ideas to incubate, which is often an opportunity lacking in many experimental creativity research studies. The reason for this is the desire to control for the influences of other individuals ideas on the participant's own work. To mitigate this influence, we ask that you do not discuss your ideas with other prior to completion within the week. In this way, artwork can only be attributed to your own efforts, which is exactly what is required for this study.

Unless otherwise stated by a participant, the default font will be Gill Sans, and the size of the type will be selected to fit an A4 sheet (allowing for a 20 mm border).

* It may be that email is unsuitable for some designs due to their complexity. In such cases, a Skype call will be arranged, and share screen enabled to allow a participant to see a live version of their design in Adobe Illustrator. For most technical support, I will undertake the session to guard against offering design suggestions, but where a level of skill is required beyond my own Adobe Illustrator support, I will engage a professional graphic designer expert in Adobe Illustrator; I will still oversee these support session to guard against advice been given.

Figure 28: Information on Technical Support for the Type Task

7.2.2.1 1st Pilot

The first pilot participant considered themselves to be a graphic design novice and had not engaged either formally, through education, or as a hobby in other domains related to art, design or craft. They completed the Image Task in the week allowed, and did not need any technical support for the task. They used Google to gather their two images and used MS Word to crop and position these on an A4 sheet of paper.

For their 2nd task (the Type Task) they did require technical support. They sent a Word document with the word they had chosen, and instructions on what they wanted.

Technically they did not know how to achieve the look they wanted in Word (indeed, it was not possible to do so) and after a few attempts asked for support. Using a feature in Adobe Illustrator, the effect requested was easy to achieve, and a copy was sent back to them, which met with their approval.

7.2.2.2 2nd Pilot

The 2nd participant also considered themselves to be a graphic design novice.

Like the first participant, they gather their images from searching on Google images. In contrast, however, the low resolution of the images when scaled up to 10 cm by 10 cm (as was the default given in the image brief), was a concern -but not one raised by the participant. The consideration of image quality was something that had not been discussed in the technical support instructions, and it was decided to amend that document to include information on both where to gather images, and the quality of image resolution that was preferred (see Appendix D). Like the first participant, both images were placed within an Adobe Illustrator template, with each image increased/decreased in size to either a 10 cm height or 10 cm width. A 10 cm by 10 cm clipping mask was then used to crop each image, and each image was placed alongside the other in the centre of an A4 sheet.

When participant two undertook the Type Task, the wording: "You are asked to choose a word and then typographically illustrate its meaning" was questioned, as it was possibly using more domain-specific language than required for those unfamiliar with graphic design. This wording was simplified, and amended to: "You are asked to choose a word and then visually communicate its meaning through the use of type."

7.2.2.3 3rd Pilot

Participant three was also a novice. For the Image Task, they supplied two images in a word file. These were added to Adobe Illustrator using the layout described in the technical support instructions and sent back. On review, the participant requested that one of the images be cropped further.

With the Type Task, participant three sent a photograph of a hand drawn sketch. Most of the design features could be identified but due to the angle of the photograph some aspects of the sketch were slightly ambiguous. A first version was created in Adobe Illustrator using the original sketch as an underlay. After receiving a digital artwork, participant three clarified a feature with a further sketch and photograph, this time, photographed directly above the sketch. The digital artwork was amended and met with approval.

After the third pilot it was considered a position of saturation had been reached; enough refinement of the wording for each creativity task and instruction for technical support had been adapted to be appropriate for novice participants, and little benefit was to be gained from continuing to pilot the creativity tasks. Of most importance, the application of technical support did not appear to influence the design a novice participant had visualised. Through photographs of sketches, and written instructions novice participants were more than capable of expressing their creative vision, and ask for amendments until the design met with their satisfaction.

7.2.3 Visualisation and Written Comprehension

As detailed in the discussion of Study 2, for practical and ethical reasons, within a single study, it was considered a maximum of two competencies could be measured before participant fatigue and attrition rate was likely to impact on findings. As each of the nine narrow abilities highlighted in Study 2 appeared to cluster around three broad themes related to creativity, language, and visualisation, these broad themes became an area of focus for Study 5.

Equally, other factors suggested this may be a useful direction for empirical study. For instance, the measure of graphic design creativity in Study 5, like Study 4, was composed of two tasks (one image only, the other text only). The opportunity to isolate these two defining features of graphic design, and test their relationship to cognitive measures of visualisation and written comprehension was novel and had a symmetry of research design. In addition, the abilities of visualisation and written comprehension appeared important to BTEC, NASAD and O*NET's graphic design competency models as a whole, and O*NET's Graphic Design job analysis placed importance on Oral Expression, Written Comprehension, Written Expression, Oral Comprehension, and Visualisation.

7.2.4 The Influence of General Intelligence

Given Mackie's findings of a significant relationship between linguistic ability and creativity within graphic design and the importance O*NET placed on reading and writing abilities, it was possible such findings could have been influenced by the broader issue of general intelligence (IQ). Thus for study 5, it was prudent to consider IQ as part of the research design (it is important to emphasise that the assessment of IQ was considered as a compounding factor rather than a focus per se).

In order to measure general intelligence, initially, the Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II) was chosen as the two subset form can be completed in about 15 minutes. These scores can give a full-scale estimate of general cognitive abilities, and, moreover, the two subtests: Vocabulary and Matrix Reasoning, fit the focus on visualisation and written comprehension in this study.

Unfortunately, upon application to Pearson, it became clear that the level of qualifications and clinical experience required to access this test was not available to the principle

investigator or the supervisory team. For this reason, a suitable well-regarded measure of IQ was required: one that did not take much longer than the WASI-II, was suitable for adults, and would be available for purchase by the principle investigator.

The Kaufman Brief Intelligence Test Second Edition (KBIT-2) matched these criteria. It was considered a well-established measure of general intelligence, was shown to be highly reliable and, through its norm-referencing of standardised scores, a comparable measure to other intelligence tests. Although KBIT-2 does not have a set time, Pearson suggests the administration took in the region of 20 minutes. A further feature was specific measurement of both verbal and non-verbal abilities; again, this was considered an asset for this study.

7.2.5 An Appropriate Size of Group

While participants were stratified between novices, intermediates and experts, to offer a range of graphic design creativity, it was important they be considered as one group, as a whole, for data analysis. In this context, thirty participants are considered to be a threshold for the use of parametric statistics (Gall, Borg & Gall, 1996). Granted, this is a topic of debate; some would prefer to set a baseline of 50 participants. For this reason, a power analysis (Figure 29) was performed and highlighted that with a sample size of 32 participants, a one-tailed correlational analysis would have 80% power to detect a moderate correlation if it was present in the data.

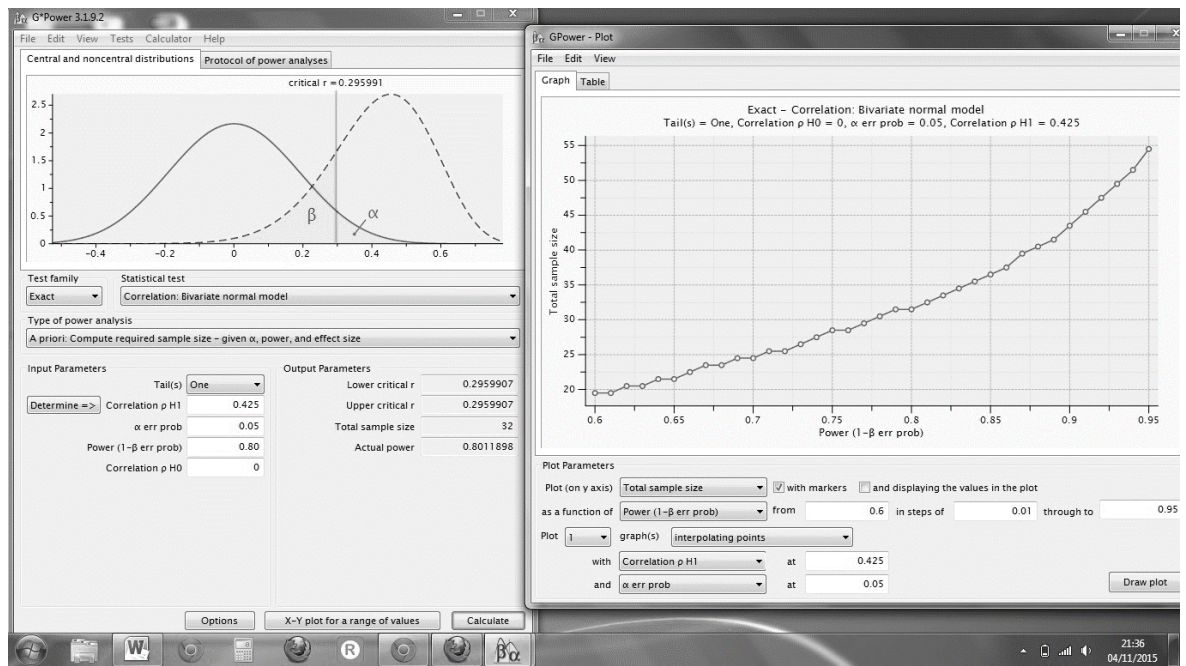


Figure 29: Screen Shot of G-Power analysis for a one-tailed correlational design

7.2.6 Main Research Question and Subordinates

Given the background described above, and based on the findings of the previous studies in this thesis, Study 5 was designed to explore the relationship between two distinct narrow abilities: visualisation (the ability to form mental images of how patterns or objects would look after certain changes, such as unfolding or rotation), and written comprehension (the ability to read and understand the meaning of words, phrases, sentences and paragraphs). Thus the original guiding research question was rephrased for Study 5 as:

To what extent are abilities, like visualisation and written comprehension, correlated with type and image graphic design creativity, when discounting for verbal and visual IQ?

7.2.6.1 Sub-Research Questions

As the dependent variable and key measurement for this study was creativity, without the robustness of how this construct was measured, all other findings for this research would

be brought into doubt. Equally, this is a concern relevant to the validity of previous studies and findings within design and creativity research. Regardless of the statistical power, significance and correlation, if the operational definition and protocols were in doubt (especially in a study where creativity is the dependent measure), then findings are open to greater criticism. For this reason, three sub-questions related to the measurement of graphic design creativity were:

- *Can the Type Task, and Image Task, remain reliable measures of graphic design creativity without pre-selection of artwork?*
- *To what extent are the Type Task and Image Task correlated with each other?*
- *To what extent are technical execution and aesthetic appeal correlated with graphic design creativity?*

7.3 Method

7.3.1 Broad Research Design

Two groups of participants were required for this study: judges and participants. For practical reasons, participants and judges were recruited that worked or live within a 30-mile radius of UCLan's Preston Campus. This enabled a suitable catchment area to include Manchester, Liverpool, Blackpool and Blackburn. Thirty-two participants undertook a battery of visualisation, written comprehension, general intelligence tests, and graphic design creativity tasks. Each participant was stratified as either an expert, intermediate or novice in terms of their experience within graphic design. After agreeing to take part, a creative Image Task was given in the first week; in the second week a creative

typographic task; a range of written comprehension and visualisation tasks, and an IQ test, in the third week. The creativity tasks were completed by participants in their own time, and unsupervised by the researcher; the third weeks' tasks were supervised by the researcher. The creative task artwork was judged by eight professional graphic designers using the Consensual Assessment Technique (CAT), and a range of statistical analysis was used as required, for example, Cronbach's alpha, Factor Analysis, and Pearson's Correlations coefficient.

7.3.2 Identification of Experts, Novices and Intermediates

Precise definitions and a number of distinctions and practical implications were considered to gather a range of graphic design expertise. With both expert and intermediate graphic designers, the recruitment procedures in many respects qualified the level of a participant's expertise. In essence, by contacting graphic design agencies, most participants from this sample were likely to be experts and, similarly, by contacting graphic design students, most were likely to be intermediates.

This was not as obvious for the novice graphic designers, where the recruitment procedure would begin through email via the OU Virtual Participant Panel database. It was quite possible that a participant could be an intermediate or expert graphic designer, though less likely given the nature of this database for psychology research.

On closer consideration, it was also feasible that an expert might be gaining a qualification by returning to academic study, and thus not strictly an intermediate purely due to their enrolment on a graphic design degree. The answer in each of these cases was to undertake a screening questionnaire prior to participation in the study.

As a basis for definition, experts, intermediates and novices were defined as follows.

Experts were classed as participants who work full-time as a graphic designer, and additionally had more than five years full-time experience as a professional graphic designer. Intermediates were classed as participants still undertaking formal training as a graphic designer, specifically students who were studying a full-time BA (Hons) degree in graphic design, with more than two years full-time study experience on the course.

Novices were classed as participants who have not undertaken formal training in graphic design, nor have been employed as a graphic designer. Each of these three levels could be identified by asking participants to judge themselves against the following statements:

Expert: I have worked as a full-time graphic designer for more than five years

Intermediate: I have studied on a full-time graphic design degree for more than two years

Novice: I have not worked as a graphic designer nor studied graphic design.

On reflection, asking participants to self-labelling as novices (or for that matter expert or intermediate) seemed unnecessary. For research sampling such distinctions are a useful shorthand, but what was required was to know the details of a participant's engagement with professional graphic design, specifically, and design, art or craft more broadly, and then decide which distinction was most appropriate (expert, intermediate, or novice) to achieve a diverse sample. As a result, each level of expertise was identified by gaining information from novice participants on the following questions, and presented as follows, via email (Figure 30):

For this study we require a range of expertise in graphic design, and I would like you to consider which statement best reflects your experience in graphic design specifically, and design, art or craft more broadly:

- 1: I work as a full-time graphic designer, and have done so for more than five years
- 2: I have worked as a designer, artist or craftsperson in the last ten years.
- 3: I am currently studying a full-time graphic design degree, and have done so for more than two years
- 4: I have studied design, art or craft either formally or as a hobby in the last ten years.
- 5: I have not studied design, art or craft either formally, or as a hobby, since leaving secondary school.

If you could return this email with the number of the statement or statements that best reflects your experience to date, that would be great. I'll then send you the first creativity tasks with instructions on how to take part.

Figure 30: Request for information of level of experience in graphic design

A consideration in the wording above was the possibility of novices with design or artistic backgrounds, but not in graphic design specifically. This could be a troublesome sampling issue. Whilst all novices may consider their expertise in graphic design to be negligible, some may have engaged, or be engaged, in tasks that could underpin graphic design creativity. A person with an interest in entering local painting competitions, for instance, may not consider themselves anything other than a novice of graphic design, but such a hobby was likely to set them apart from novices who have no interest in visual culture or participate in "artistic" pursuits. For this reason, the question 4 was added to the screening questions.

7.3.3 Participants

In all cases, participants were emailed a copy of the information sheet (Appendix E), and a follow-up phone call was made two to three days after this initial email contact.

The recruitment of participants with no graphic design experience was sought via the OU Virtual Participant Panel (VPP). Access to the VPP is a vetted application process for OU members of staff, or OU research students, to aid with the recruitment of research studies in psychology; it is only available to studies that have received ethical approval. Once access was granted, a VPP database search was restricted to specific postcode locations within a 30-mile radius of the University of Central Lancashire. This database had around 1500 volunteers willing to undertake research studies.

Graphic designer Intermediates were recruited from graphic design students studying on the BA (Hons) Graphic Design course at UCLan. These students had completed the second year of their course. Initially, access to this year group was sought through consultation with the graphic design staff at UCLan.

Contact details were acquired for professional graphic designers within a 30-mile radius of UCLan's Preston Campus. Much of this information was freely available via the web, for example via yell.com results 67 graphic designers were within a ten-mile radius of PR1 2HE.

In total 118, professional graphic designers, student graphic designers and members of the general public were contacted. 32 took part and completed all the task; 18 females and 14 males, with a mean age of 32.69 years ($SD = 11.27$), that ranged from 19 to 58 years of age. Specific sampling details for novices, intermediates and experts are as follows.

7.3.3.1 Novices

39 novices were contacted, of which 14 agreed to take part. Three withdrew after receiving the emailed instructions for the first task (image juxtaposition), and 11 continue to complete all tasks. 10 females and 1 male took part. The mean age was 35.27 ($SD = 7.95$ years) and ranged from 23 to 49. Seven of the group had not studied design, art or craft either formally, or as a hobby, since leaving secondary school. Four of the group said they had studied design, art or craft either formally or as a hobby in the last ten years. Example comments for this rating were: "I do a lot of crafts as a hobby. Crochet, jewellery" or, "I have done a little silversmithing and bookbinding as a hobby - more technical than design, but that perhaps makes me a 4? No formal study, or art or graphics at all."

7.3.3.2 Intermediates

39 intermediates, of which 12 agreed to take part. Two withdrew after receiving the emailed instructions for the first task, 10 continued to complete all tasks. Six females, and four males. The mean age was 20.70 years ($SD = 1.34$) and ranged from 19 to 24. All described themselves as currently studying a full-time graphic design degree and had done so for more than two years.

7.3.3.3 Experts

40 experts were contacted, of which 12 agreed to take part; one dropped out after receiving the emailed instructions for the first task, and 11 continued to complete all tasks. Two females, and nine males. Years of experience in graphic design ranged from 4 to 35 years, with only one designer having less than 5 years' experience; the mean experience was 19 years ($SD = 9.77$ years). The mean age was 41 ($SD = 10.08$ years) and ranged from 26 to 58 years. The majority of experts described themselves as a graphic designer, designer or creative; other titles were art director, creative director, and brand engagement specialist.

7.3.4 Measures

The creative Image Task and creative typographic task were the same as those used in Study 4. As described above in the background pilot results, minor adaptations were made to accommodate novice participants.

Based on the findings of Study 2, three measures of visualisation were used: Marks' (1972) Vividness of Visual Imagery Questionnaire (VVIQ); the Purdue spatial visualisation test, and from the Educational Testing Services' (ETS) the Snowy Pictures test (CS-3). As detailed in section 4.5.3 each test had long-standing use as measures of visualisation.

Likewise, based on findings of Study 2, two measures of written comprehension were used. The Nelson-Denny Reading Test (form G) was chosen as the main measure of written comprehension. It was a test relevant for adults and college students, took 20 minutes to complete, a long-standing history and had remained in current use as a measure of written comprehension. To compliment this, the ETS Advanced Vocabulary Test was chosen as the secondary measure of written comprehension.

As both written/reading comprehension and visualisation are classed as narrow abilities with CHC theory, one of the main covariants that could impact on this research design was the extent to which these narrow abilities are explained by higher order abilities and general intelligence. To account for this Kaufman's Brief Intelligence Test (version 2) was used as a measure of general intelligence, as described in section 7.2.4 (The influence of general intelligence)

7.3.4.1 Order of Tasks

For a number of reasons, the order participants would undertake tasks were staged in a particular sequence: with the creativity tasks first, followed by the psychometric/timed measures used for this study. Much of this rationale was for practical considerations due to the possibility of non-completion. For example, the creativity task could be completed in a participant's own time; if, however, they did not complete the tasks, then little research time was lost as a result. In contrast, the battery of tests used to measure visualisation, written comprehension, and IQ all required a meeting to take place, in addition to the financial costs associated with using most of these tests. Should a participant leave the study, it was less disruptive for this to occur after the creativity tasks, rather than after the psychometrics.

Furthermore, it was felt that the psychometric tasks were likely to be a more burdensome

experience for creative intermediates and professionals, and less engaging than the creativity tasks. Thus, the creativity tasks offered an appealing way to build participants interest in completing the study.

Of the creativity task, the Image Task appeared to be the less technically demanding for participants (particularly those with no graphic design background), and would allow a staged approach to the technical support that was available to them. For this reason, the Image Task was the first task that all participants would undertake, followed by the Type Task. Upon completion of the creativity tasks, the arrangement would be made to meet and undertake the remaining measures used in this study. In practice, arrangements to meet were discussed at the start of a participant's involvement in the study, to both act as an impetus for the completion of the creativity tasks, and to allow diaries to be scheduled in advance.

In relation to the tasks to measure visualisation, written comprehension and general intelligence, both order effect and fatigue effect needed to be considered relative to placing participants at ease. A non-timed test was considered the best route to begin the testing session: the choice was between the KBIT and the Marks' VVIQ. Given this was an introductory task, Marks was chosen, as this was likely to take less time, and was a task that required minimal instructions, with minimal test like condition associated with a right or wrong answer, as was the case with the KBIT. Although Marks is not a timed test it was expected that most participants would complete within 5 to 15 minutes.

Starting with Marks set the first focus on visualisation, and the follow-on test in this battery was the ETS Snowy Pictures. At this point, participants were introduced to a 3-minute countdown application placed on the table in front of them. The Snowy Pictures had two

parts each of 3 minutes duration. The last test for visualisation was the Revised PSVT:R (a time test of 20 minutes).

The visualisation tasks were followed by written comprehension, starting with the ETS Advanced vocab test. Like the ETS snowy pictures, the test is in two parts of 4 minutes each. This was followed by the Nelson-Denny Test, written comprehension component of 20 minutes.

The final task was the KBIT-2, which has three distinct sections: verbal knowledge, matrices, riddles. It is not a timed task, but the expectation was that most adults could complete the task with 20 -25 minutes.

After completion of the KBIT, participants had the chance to ask any questions and were briefed on the outcomes of the study, and timescales for results.

7.3.5 Judges

Eight professional graphic designers were recruited to rate the artworks participants create, and each designer had at least 5 years professional experience in graphic design. Of the 40 experts contacted above, eight agreed to take part as judges. The age range for judges was between 31 and 55; the mean was 41.63 years (SD = 8.88 years). Four females and four males. Years of experience in professional graphic design ranged from 9 to 34 years; the mean was 19.13 years, (SD = 8.43 years)

7.3.5.1 Judges Ratings Procedures

It was explained to each judge they would be asked to assess two briefs (one focused on the creative use of image, the other focused on the creative use of type), and that judges would rate these 32 artworks for creativity, amongst some other measures (but, no

mention of technical execution or aesthetic appeal was given at this stage).

Also highlighted was that the study had 32 participants, with a range of graphic design expertise: going from novices who have no experience of graphic design, through to year 2 students of graphic design, through to professional graphic designers with over 5 years full-time experience (many had much more, 15 to 20 years). As some of the participants had no experience of software like Adobe Illustrator, these participants had been offered technical support.

It was explained to judges they would use the laptop to view the artwork, and that a thumbnail of the images would allow them to rate the artwork through sorting these onto the rating sheet provided.

The first brief was given to the judges, and they were told this was the exact same brief seen by the participants, and that at the back was the instruction on technical support that was provided to all participants. Judges were then left to read these documents, and the researcher set up the laptop to view the artwork (this was shown in full-screen mode).

Specifically, the creativity only instructions were worded as follows:

Instructions for Judges: How to rate these artworks

There is only one criterion in rating these artworks: creativity. We realise that creativity probably overlaps other criteria one might consider (for example: aesthetic appeal, or technical execution) but we ask you to rate the artworks solely on the basis of their creativity. There is no need to explain or defend your ratings in any way; we ask only that you use your own sense of which is more or less creative (**relative to the other artworks provided**).

Please look through these artworks three times, and rate them for creativity.

- The first time familiarise yourself with all the artworks provided.
- The second time, group the artworks into Low, Medium, or High ratings.

- The third time, assign a numerical rating between 1 and 6 (1's being the least creative and 6's being the most creative).

There should be a roughly even number of artworks at each of the six levels. It is very important that you use the full 1-6 scale.

The instructions used the exact same wording as Study 4's Baer's instructions. In contrast to Study 4, in this study, once judges had read the instructions, a number of points were verbally re-emphasized within the instructions by the researcher. First, that they were to rate the artwork on creativity only (without considering technical execution and aesthetic appeal). Second that their ratings were relative to the artworks provided, rather than an external standard of creativity (such as award-winning creativity). Third, that it was important to have an even number of artworks rated in each of the 1 to 6 categories, this need not be strictly so, but roughly equal.

A few other points were made that were not in the instruction but seemed to be helpful for judges. As the artwork was in both landscape and portrait, it was useful to arrange the thumbnails to reflect this. Additionally, whilst the researcher was happy to answer questions at any point during the process, if those questions related to artwork such as, "What does this say?" or "What is this image?" then the researcher would not give details on this, as it was likely to influence judges rating and bias their assessment. That said, any question about what a word said, or a particular image would be discussed at the end of the judge's assessment.

With this in place, and rapport established, the judges were given the laptop, shown how to navigate through the artwork, and had the rating sheet placed in front of them with the thumbnails stacked and in the "medium" area of the rating sheet. The rating sheet was the exact same format as that described in Study 4, Section 6.3.6 Procedures, and illustrated

in Figure 23. The researcher then left the judge to undertake their assessment without supervision.

Once the creativity only assessment was completed, a post-it note was placed on the rating sheet to denote this was a creativity only assessment and a photograph was taken of the sorted artwork. As a back-up, two separate cameras were used to take a photograph.

Judges were then given the instructions for either technical execution or aesthetic appeal, and the process above was repeated. The wording of instruction for rating aesthetic appeal (see below) was virtually identical to that of technical executions: the only change was to replace the phrase aesthetic appeal with that of technical execution.

Aesthetic Appeal

Please look through these artworks and rate them for aesthetic appeal. There is no need to explain or defend your ratings in any way; we ask only that you use your own sense of which is more or less aesthetically appealing **(relative to the other artworks provided)**.

- Familiarise yourself with all the artworks provided.
- Group the artworks into Low, Medium, or High ratings.
- Assign a numerical rating between 1 and 6 (1's being the least aesthetically appealing and 6's being the most aesthetically appealing).

There should be a roughly even number of artworks at each of the six levels. It is very important that you use the full 1-6 scale.

Judges read these new instructions, in this example for assessing aesthetic appeal, and after this, it was pointed out that the details relevant to their assessment were more likely to be found through viewing the digital artwork on the laptop rather than looking at the thumbnails. The thumbnail artwork was randomly re-sorted and placed back to the top of the rating sheet. Judges were then left to begin their assessment.

Once this 2nd assessment was completed, the same procedure was followed but for the remaining instructions, either aesthetic appeal or technical execution. In short: judges were given a new set of instructions; allowed time to read these; the research then re-emphasizes details to be found on the screen rather than in the thumbnails; the thumbnails were randomly rearranged with artwork to the top of the rating sheet; judges were left to begin their assessment.

Exactly the same procedures were followed for the second task. In which order judges undertook either the image or Type Task, and in which order they undertook the assessment of technical or aesthetic appeal was determined by the randomization of judges into counterbalanced conditions (Table 31).

Table 31: Random allocation of judges into counterbalanced conditions

Judge	Artwork	1st Brief	1st Order after C	2nd Order after C	2nd Brief	1st Order after C	2nd Order after C
2	Set A	Image	Aesthetic	Technical	Type	Technical	Aesthetic
4	Set B	Image	Aesthetic	Technical	Type	Technical	Aesthetic
6	Set C	Image	Technical	Aesthetic	Type	Aesthetic	Technical
8	Set D	Image	Technical	Aesthetic	Type	Aesthetic	Technical
1	Set E	Type	Aesthetic	Technical	Image	Technical	Aesthetic
5	Set F	Type	Aesthetic	Technical	Image	Technical	Aesthetic
7	Set G	Type	Technical	Aesthetic	Image	Aesthetic	Technical
3	Set H	Type	Technical	Aesthetic	Image	Aesthetic	Technical

7.3.6 Ethics, Data Protection, Location and Recompense

Study 5 followed the same guidelines as described in section 6.3.2 of Study 4, and gained ethical approval from the OU Human Research Ethics Committee.

Consent was sought via participants signing the consent forms (Appendix F). Any

participant who wished to withdraw from the study was informed they could contact the principal investigator (contact details were provided on both the consent form and information sheet).

Participants, through the information sheet, were made aware they could withdraw up to the point of judges assessing the artwork. Their artwork and test results gathered up to the period of withdrawal would be used, if they were happy for this to occur. Otherwise, they could request the images of their artwork and test results collected by this study to be destroyed and no further use made of them.

In the case of professional graphic designer judges, through the information sheet, they were made aware they could withdraw up to the point of data analysis. Their ratings gathered up to the period of withdrawal may be used, if they were happy for this to occur. Otherwise, they could request their ratings be destroyed and no further use made of them.

As with Study 4, the research mitigated the potential to reinforced self-labelling and issues of social standing through not using illustrations of any artwork provided. Thus, it was not possible for a participant taking part to know how judges rated their artwork from the results of the study.

All artwork collected receive an anonymous code at the start of the research. Only the principle investigator knew which codes relate to which participants. This information was stored in line with the DPA guidelines (see section 6.3.2) and was fully anonymised.

As detailed previously, prior to undertaking the creativity tasks, the level of experience in graphic design was checked through a short questionnaire for each participant, after they

had been emailed a consent form and information sheet. Participants undertook graphic design creativity tasks in their own time and at a location of their choosing, and emailed finished artwork to the principle investigator. The location of other tasks changed depending on the participants. Graphic design novice participants recruited via the OU Virtual Participant Panel attended either UCLan's Preston campus or a home visit was arranged. Graphic design intermediate participants on the BA (Hons) Graphic Design at UCLan attended on campus. Graphic design expert participants arranged to meet at their design studios or attended on campus.

The majority of judges assessed artwork at their own premises, as this was a convenient location for judges, without placing more demands on their time through travel. However, as with Study 3 and 4, some judges preferred to attend the University, and this took place in the Principle Investigators office within the School of Art, Design and Fashion.

If a participant chose to travel to the Preston UCLan campus, their travel expenses were recompensed. Upon completion of the tasks for this study, each participant was given a £10 Amazon voucher as a token of thanks. They were not aware of this prior to, or during, engagement with the study. As participants had undertaken in the region of over 3-4 hours of participation, it did not cover the time costs they had incurred but was a gesture of thanks for participation. Funding was sought, and granted, by UCLan's School of Art, Design & Performance to cover the cost of the written comprehension and visualisation tests; the travel expense cost of participants, and for the small token of appreciation for participation in the study.

7.4 Results

The following section details the results for the dependent variable of creativity, the independent variables of visualisation, written comprehension and the compounding

variable general intelligence. With the assumption of normality of distribution checked, the correlation coefficient is presented.

As multiple correlations would be explored in study 5, the possibility of finding a significant correlation by chance, when one did not exist, was considered (also known as a type I error). In these instances, a popular and frequently adjustment applied by researchers is the Bonferroni correction. The purpose being to raise the threshold required for identifying statistical significance relative to the number of tests undertaken. However, raising the threshold for a type I error has to be considered in relation to the risk of a Type II error (i.e. not finding statistical significance when it does exist). Armstrong (2014) in his systematic review of the Bonferroni correction was critical of the ubiquitous application of the technique. His conclusion identified several contexts in which Bonferroni was not useful, and should not be applied, of which three are directly relevant to Study 5. Firstly, where the nature of the research is exploratory it is important researchers do not overlook correlations. For example, where the intention is to explore correlations which would then form the basis for further more definitive testing should significant relationships be found. Secondly, and related to above, when the researcher wants to avoid a type II error. Thirdly, where multiple use will be made of Pearson's r . Each of these instances is relevant to the context of study 5 and for these reasons the Bonferroni correction was not applied.

7.4.1 Inter-Rater Reliability Scores for Image and Type Tasks Creativity

For the Image Task, the Cronbach's alpha was computed for each assessment, with the following results: Creativity only was 0.81; Aesthetic Appeal was 0.75; Technical execution was 0.7. For the Type Task, Creativity only was 0.92; Aesthetic Appeal was 0.88; Technical execution: Cronbach's alpha of 0.88. Each score was above the 0.7 alpha

level commonly used as a threshold for an appropriate level of consensus among judges.

7.4.2 Normality of Distribution

7.4.2.1 Image Task

The combined scores for each of the three image ratings were computed and checked for normality of distribution; histograms were plotted for each rating. In addition, to check data would satisfied assumption of skewness for a correlations analysis, skewness and standard error of skew was computed. A general guide is that a skew statistic of ± 3.0 is an indicator of significant skew. These results, and range, mean and standard deviation are shown below (Table 32).

Table 32: Descriptive statistics for calculating Z-skew of Image Task

	N	Min.	Max.	Mean	SD	Skew	SE	Kurtosis	SE
Image_Creativity	32	9	45	28.16	9.284	.093	.414	-.236	.809
Image_Aesthetic	32	11	40	27.72	8.085	-.186	.414	-.982	.809
Image_Technical	32	14	48	28.31	8.244	.492	.414	-.043	.809

In each rating, skew and standard error of skew was converted to a Z-score:

Image_Creativity gave a Z-skew of 0.22; Image_Aesthetic, -0.45; Image_Technical, 1.19.

All Z-skew statistics were within the threshold of ± 1.96 as suggested by Cramer & Howitt (2004) suggesting the data was appropriate for parametric analysis. On this basis,

Pearson's r was used to calculate the correlation coefficient between each rating:

7.4.2.2 Type Task

Likewise for the Type Task combined scores for each of the three ratings were computed and checked for normality of distribution, histograms were plotted for each rating. To check data would satisfy assumptions of skewness for correlation coefficient analysis, skewness and standard error of skew were computed. These results, and range, mean

and standard deviation are shown below (Table 33):

Table 33: Descriptive statistics for calculating Z-skew of Type Task

	N	Min.	Max.	Mean	SD	Skew	SE	Kurtosis	SE
Type_Creativity	32	10	47	27.78	11.152	.274	.414	-1.166	.809
Type_Aesthetic	32	8	46	27.03	10.149	.150	.414	-.668	.809
Type_Technical	32	8	44	27.37	10.232	-.164	.414	-.734	.809

In each rating, skew and standard error of skew was converted to a Z-score:

type_Creativity gave a Z-skew of 0.66; Type_Aesthetic, 0.36; Type_Technical, -.40. All Z-skew statistics were within the threshold of +/-1.96, suggesting the data was appropriate for parametric analysis.

7.4.2.3 Visualisation Tasks

To check data would satisfied assumption of skewness for correlation coefficient analysis, skewness and standard error of skew was computed for all visualisation measures: Marks' VVIQ, ETS Snowy Pictures and Purdue PSVT:R. In each rating, skew and standard error of skew (Table 34) was converted to a Z-score: VVIQ gave a Z-skew of -0.18; ETS Snowy Picture, -1.11; PSVT:R, -0.57. All Z-skew statistics were within the threshold of +/-1.96, suggesting each of these measures of visualisation was appropriate for parametric analysis.

Table 34: Descriptive statistics for calculating Z-skew for visualisation measures

	N	Min.	Max.	Mean	SD	Skew	SE	Kurtosis	SE
VVIQ	32	37	80	59.09	10.325	-.073	.414	.245	.809
ETS Snowy	32	6	19	13.47	3.142	-.458	.414	.119	.809
PSVT:R	32	7	26	17.38	5.191	-.237	.414	-.890	.809

7.4.2.4 Written Comprehension Tasks

The ETS advance vocabulary Test states in the instructions to participants that: "Your score will be the number of marked correctly minus a fraction of the number marked

incorrectly". In the scoring manual, ETS instructions are not specific on how to calculate this fraction, thus, for this study were interpreted to be the total wrong answers divided by correct answers. This is smaller for those that have fewer wrong answers relative to correct answers, and larger for those that have more wrong answers to correct answers. For example, a participant who get 15 correct answers and 10 incorrect would be deducted 0.67, thus a score of 14.3. A participant who get 10 correct and 15 incorrect would be deducted 1.5, thus a score of 8.5. This seemed the most reasonable way to interpret the ETS instruction and the statement to participants that "it will not be in your advantage to guess".

Form g of the Nelson-Denny was used for the written comprehension component.

Following the guidance described in the *Manual for Scoring and Interpretation* (Brown, Fishco & Hanna, 1993) raw scores were converted to the standard scale scores for both reading rate and comprehension. Standard scales were recommended for statistical analysis given they had equal interval characteristics and based on the normalisation of a number of reference studies that pooled a variety of educational populations (for example, university students at various stages of their degree). Standardised scale norms have a mean of 200 and a standard deviation of 25.

To check data would satisfy the assumption of skewness for correlation coefficient analysis, skewness and standard error of skew were computed (Table 35). These results, and range, mean and standard deviation are shown below:

Table 35: Descriptive statistics for calculating Z-skew of written comprehension measures

	N	Min.	Max.	Mean	SD	Skew	SE	Kurtosis	SE
ETS A. Vocab	32	-8	36	17.11	8.562	-.168	.414	1.643	.809
N-D RRate	32	171	286	214.09	26.255	.816	.414	.961	.809
N-D RComp	32	182	248	218.22	17.974	-.480	.414	-.632	.809

In each rating, skew and standard error of skew was converted to a Z-score: ETS Advanced Vocabulary gave a Z-skew of -0.41.; Nelson-Denny Reading Rate, 1.97; Nelson-Denny written comprehension, -1.16. Z-skew statistics were within the threshold of ± 1.96 for all measures with the exception of the Nelson-Denny Reading Rate.

A histogram was plotted for the Nelson-Denny Reading Rate measures (Figure 31): it was interpreted that the Nelson-Denny Reading Rate showed enough of a normal distribution to proceed, in addition, reading rate was not a key feature for why the Nelson-Denny was chosen; written comprehension was, and in this respect, the data was within an appropriate threshold for parametric analysis.

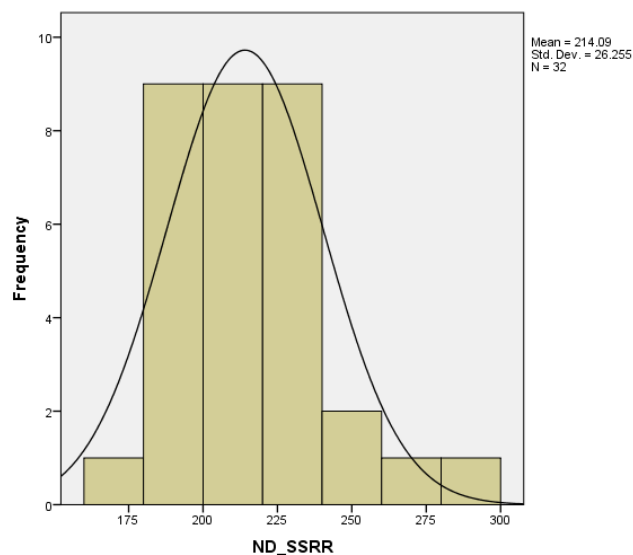


Figure 31: Histogram of Nelson-Denny Reading Rate scores

7.4.2.5 General Intelligence

Raw scores for the KBIT were standardised following the procedures identified in their respective scoring manual. To check data would satisfy the assumption of skewness for correlation coefficient analysis, skewness and standard error of skew was computed (Table 36). These result, and range, mean and standard deviation are shown below:

Table 36: Descriptive statistics for calculating Z-skew of general intelligence scores

	N	Min.	Max.	Mean	SD	Skew	SE	Kurtosis	SE
KBIT_Verbal	32	89	135	105.84	11.623	.665	.414	-.026	.809
KBIT_N-Verbal	32	51	130	104.16	15.423	-1.079	.414	3.561	.809
KBIT_Com	32	66	129	105.53	12.776	-.603	.414	1.697	.809

In each rating, skew and standard error of skew was converted to a Z-score: KBIT Verbal gave a Z-skew of 1.61; KBIT Non-Verbal, -2.61; KBIT Composite score, -1.46. Z-skew statistics were within the threshold of ± 1.96 for all measures with the exception of the KBIT Non-Verbal scores. Histograms were plotted for the KBIT Non-Verbal measure (Figure 32).

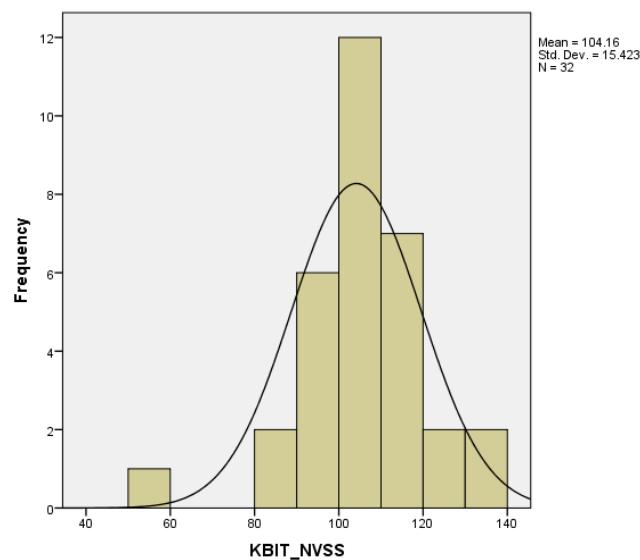


Figure 32: Histogram of KBIT Non-Verbal scores

7.4.3 Correlations Between Creativity, Aesthetic Appeal and Technical Execution

Pearson's r was used to calculate the correlation coefficient between each rating and task (Table 37).

Table 37: Guidance on the strength of relationship associated with Pearson's r .

+.70 or higher	Very strong positive relationship
+.40 to +.69	Strong positive relationship
+.30 to +.39	Moderate positive relationship

+ .20 to + .29 weak positive relationship
 + .01 to + .19 No or negligible relationship
 - .01 to - .19 No or negligible relationship
 - .20 to - .29 weak negative relationship
 - .30 to - .39 Moderate negative relationship
 - .40 to - .69 Strong negative relationship
 - .70 or higher Very strong negative relationship

7.4.3.1 Image Task

Each correlation (Table 38) was positive at a significant level of 0.01 (1 tailed), with a very strong correlation between aesthetic appeal and technical execution ($r = 0.82$); a strong correlation between creativity and technical execution ($r = 0.68$), and strong/moderate correlation between creativity and aesthetic appeal ($r = 0.48$).

Table 38: Correlations for the Image Task

		Image_Creativity	Image_Aesthetic	Image_Technical
Image_Creativity	Pearson Correlation	1	.480**	.679**
	Sig. (1-tailed)		.003	.000
	N	32	32	32
Image_Aesthetic	Pearson Correlation	.480**	1	.819**
	Sig. (1-tailed)	.003		.000
	N	32	32	32
Image_Technical	Pearson Correlation	.679**	.819**	1
	Sig. (1-tailed)	.000	.000	
	N	32	32	32
**. Correlation is significant at the 0.01 level (1-tailed).				

7.4.3.2 Type Task

Each correlation (Table 39) was positive at a significance level of 0.01 (1 tailed), and each suggestive of very strong correlations: creativity and aesthetic appeal ($r = 0.93$); aesthetic appeal and technical execution ($r = 0.90$); creativity and technical execution ($r = 0.87$).

Table 39: Correlations for the Type Task

		Type_Creativity	Type_Aesthetic	Type_Technical
Type_Creativity	Pearson Correlation	1	.931**	.871**
	Sig. (1-tailed)		.000	.000
	N	32	32	32
Type_Aesthetic	Pearson Correlation	.931**	1	.902**
	Sig. (1-tailed)	.000		.000
	N	32	32	32
Type_Technical	Pearson Correlation	.871**	.902**	1
	Sig. (1-tailed)	.000	.000	
	N	32	32	32
**. Correlation is significant at the 0.01 level (1-tailed).				

7.4.3.3 Factor Analysis of Correlations

In addition to the correlations above, a Principal Components Analysis on the correlation coefficients for both the Type Task and Image Task was computed. The total variance, Scree plot, and component matrix are shown in Table 40 to 43, and figures 33 and 34. In each set of results, evidence can be found that a substantial single component likely to be present underpins judges' assessment of creativity, technical execution, and aesthetic appeal.

For the Type Task, only one principal component with an eigenvalue greater than one was identified, and the scree plot is characteristic of a data set with a single principal component. This single component explains 93% of the variance in the data.

Table 40: PCA of Type Task

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.803	93.430	93.430	2.803	93.430	93.430
2	.133	4.440	97.870	.133	4.440	97.870
3	.064	2.130	100.000	.064	2.130	100.000
Extraction Method: Principal Component Analysis.						

Table 41: Type Task PCA component matrix

Component Matrix ^a			
	Component		
	1	2	3
CREATIVITY	.966	-.212	.145
AESTHETIC	.977	-.073	-.199
TECHNICAL	.956	.288	.057
Extraction Method: Principal Component Analysis.			
a. 3 components extracted.			

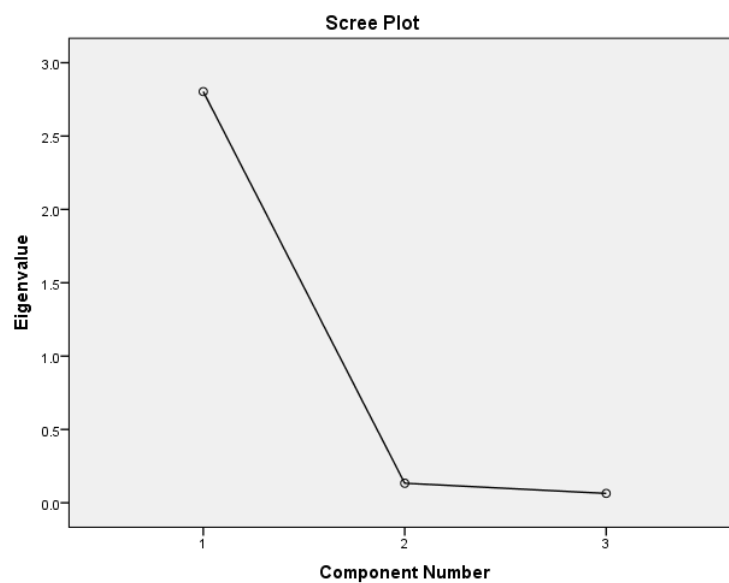


Figure 33: Scree plot for Type Task

For the Image Task, only one principal component with an eigenvalue greater than one was identified, and the scree plot is also characteristic of a data set with a single principal component. This component explains 78% of the variance in the data.

Table 42: PCA of Image Task

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.328	77.607	77.607	2.328	77.607	77.607
2	.534	17.788	95.395	.534	17.788	95.395
3	.138	4.605	100.000	.138	4.605	100.000
Extraction Method: Principal Component Analysis.						

Table 43: Image Task PCA component matrix

Component Matrix ^a			
	Component		
	1	2	3
CREATIVITY	.805	.584	.107
AESTHETIC	.879	-.428	.212
TECHNICAL	.953	-.098	-.286
Extraction Method: Principal Component Analysis.			
a. 3 components extracted.			

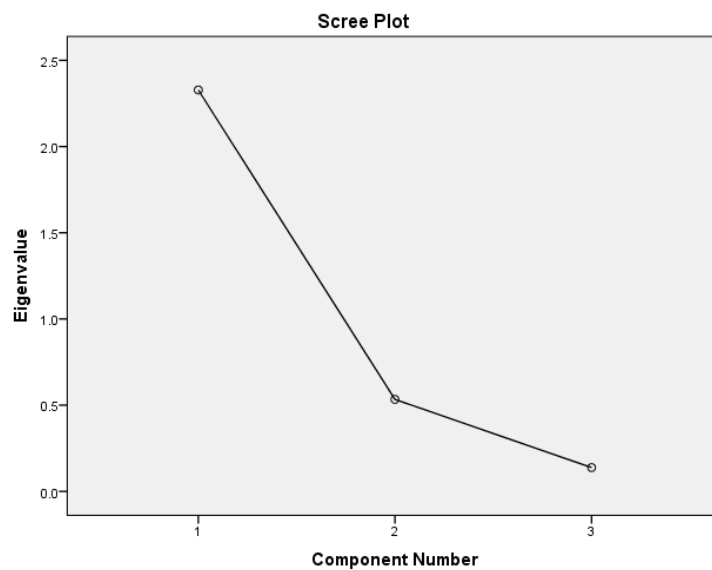


Figure 34: Scree Plot for Image Task

7.4.4 Correlations Between Type and Image Creativity

Based on the previous results from Study 4, the expectation was that a relationship would

exist between creativity on the Type Task and the Image Task. The correlation between type creativity and image creativity in this study was significant ($p < .05$), and showed a moderate positive relationship ($r = 0.32$, one-tailed).

7.4.5 Correlations Between Creativity and Visualisation

No significant correlation was found between visualisation and graphic design creativity.

Correlations were computed for Total Creativity, and separately for both the Type Task and the Image Task.

7.4.5.1 Total Creativity

The results found that the combined rating of the Image Task and Type Task showed no significant correlations with the Marks VVIQ, Snowy Pictures, or the PSVT:R. The strength of the relationship was in all cases weak to negligible (Table 44).

Table 44: Correlations between total creativity and visualisation measures

Correlations		TotalCreativity	VVIQ	ETSSnowy	PSVTR
TotalCreativity	Pearson Correlation	1	.104	.025	.238
	Sig. (1-tailed)		.286	.447	.095
	N	32	32	32	32
VVIQ	Pearson Correlation	.104	1	.109	.298*
	Sig. (1-tailed)	.286		.276	.049
	N	32	32	32	32
ETSSnowy	Pearson Correlation	.025	.109	1	-.252
	Sig. (1-tailed)	.447	.276		.082
	N	32	32	32	32
PSVTR	Pearson Correlation	.238	.298*	-.252	1
	Sig. (1-tailed)	.095	.049	.082	
	N	32	32	32	32

*. Correlation is significant at the 0.05 level (1-tailed).

7.4.5.2 Total Creativity Correlated With Visualisation When Controlling for KBIT Composite IQ Scores

After accounting for the influence of general ability, the situation remained similar to that described for Table 44. The combined rating of the Image Task and Type Task showed no significant correlations (Table 45) with the Marks VVIQ, Snowy Pictures, or the PSVT:R. The strength of the relationship remained in all these cases weak to negligible.

Table 45: Correlations controlled for general intelligence

Correlations			TotalCreativity	VVIQ	ETSSnowy	PSVTR
Control Variables						
KBIT_CompSS	TotalCreativity	Correlation	1.000	.109	.025	.245
		Significance (1-tailed)	.	.279	.447	.092
		df	0	29	29	29
	VVIQ	Correlation	.109	1.000	.208	.247
		Significance (1-tailed)	.279	.	.131	.090
		df	29	0	29	29
	ETSSnowy	Correlation	.025	.208	1.000	-.201
		Significance (1-tailed)	.447	.131	.	.139
		df	29	29	0	29
	PSVTR	Correlation	.245	.247	-.201	1.000
		Significance (1-tailed)	.092	.090	.139	.
		df	29	29	29	0

7.4.5.3 Type Task Creativity Correlated With Visualisation

Taken in isolation, no significant correlation was found between Type Creativity and Marks VVIQ, Snowy Pictures, or the PSVT:R. The strength of the relationship was in all cases weak to negligible (Table 46).

Table 46: Correlations between Type Task and visualisation measures

Correlations		Type_Creativity	VVIQ	ETSSnowy	PSVTR
Type_Creativity	Pearson Correlation	1	.021	-.130	.161
	Sig. (1-tailed)		.454	.240	.190
	N	32	32	32	32
VVIQ	Pearson Correlation	.021	1	.109	.298 [*]
	Sig. (1-tailed)	.454		.276	.049
	N	32	32	32	32
ETSSnowy	Pearson Correlation	-.130	.109	1	-.252
	Sig. (1-tailed)	.240	.276		.082
	N	32	32	32	32
PSVTR	Pearson Correlation	.161	.298 [*]	-.252	1
	Sig. (1-tailed)	.190	.049	.082	
	N	32	32	32	32

*. Correlation is significant at the 0.05 level (1-tailed).

7.4.5.4 Type Task Creativity Correlated With Visualisation When Controlling for KBIT Composite IQ Scores

After accounting for the influence of general ability, the situation remained similar to that described for Table 46. No significant correlation was found between Type Creativity and Marks VVIQ, Snowy Pictures, or the PSVT:R. when controlling for KBIT Composite scores. The strength of the relationship was in all cases weak to negligible (Table 47). In each case, it was considered that scatterplots did not show signs of a non-linear correlation to be present.

Table 47: Correlations controlled for general intelligence

Correlations			Type_Creativity	VVIQ	ETSSnowy	PSVTR
Control Variables	Type_Creativity	Correlation	1.000	.057	-.168	.193
		Significance (1-tailed)	.	.380	.184	.149
		df	0	29	29	29
	VVIQ	Correlation	.057	1.000	.208	.247
		Significance (1-tailed)	.380	.	.131	.090
		df	29	0	29	29
	ETSSnowy	Correlation	-.168	.208	1.000	-.201
		Significance (1-tailed)	.184	.131	.	.139
		df	29	29	0	29
	PSVTR	Correlation	.193	.247	-.201	1.000
		Significance (1-tailed)	.149	.090	.139	.
		df	29	29	29	0

7.4.5.5 Image Task Creativity Correlated With Visualisation Scores

No significant correlation was found between Image Creativity and Marks VVIQ, Snowy Pictures, or the PSVT:R. The strength of the relationship was in all cases weak to negligible (Table 48).

Table 48: Correlations between Image Task and visualisation measures

Correlations		Image_Creativity	VVIQ	ETSSnowy	PSVTR
Image_Creativity	Pearson Correlation	1	.160	.200	.233
	Sig. (1-tailed)		.191	.136	.100
	N	32	32	32	32
VVIQ	Pearson Correlation	.160	1	.109	.298*
	Sig. (1-tailed)	.191		.276	.049
	N	32	32	32	32
ETSSnowy	Pearson Correlation	.200	.109	1	-.252
	Sig. (1-tailed)	.136	.276		.082
	N	32	32	32	32
PSVTR	Pearson Correlation	.233	.298*	-.252	1
	Sig. (1-tailed)	.100	.049	.082	
	N	32	32	32	32

*. Correlation is significant at the 0.05 level (1-tailed).

7.4.5.6 Image Task Creativity Correlated With Visualisation When Controlling for KBIT Composite IQ Scores

After accounting for the influence of general ability (Table 49), the situation remained similar to that described for Table 48. No significant correlation was found between Image Creativity and Marks VVIQ, Snowy Pictures, or the PSVT:R. The strength of the relationship was in all cases weak to negligible. In each case, it was considered that scatterplots did not show signs of a non-linear correlation to be present.

Table 49: Correlations controlled for general intelligence

Correlations			Image_Creati vity	WVQ	ETSSnowy	PSVTR
Control Variables						
KBIT_CompSS	Image_Creativity	Correlation	1.000	.129	.247	.211
		Significance (1-tailed)	.	.245	.090	.128
		df	0	29	29	29
	WVQ	Correlation	.129	1.000	.208	.247
		Significance (1-tailed)	.245	.	.131	.090
		df	29	0	29	29
	ETSSnowy	Correlation	.247	.208	1.000	-.201
		Significance (1-tailed)	.090	.131	.	.139
		df	29	29	0	29
	PSVTR	Correlation	.211	.247	-.201	1.000
		Significance (1-tailed)	.128	.090	.139	.
		df	29	29	29	0

7.4.6 Correlations Between Creativity and Written Comprehension

For most correlations, the results of Study 5 found no significant correlations with the ETS Advanced Vocabulary Test or either reading component of the Nelson-Denny reading test (Reading Rate or Comprehension) for Total Creativity (Table 50) or the Image Task (Table 54). One significant negative correlation (Table 52) was found between the Type Task and the ETS Advanced Vocabulary Test ($p < 0.05$, $r = -0.31$, one-tailed), but after accounting for the influence of general intelligence, this was no longer significant (Table 53).

7.4.6.1 Total Creativity Correlated With Written Comprehension Scores

The results of Study 5 found that the combined rating of the Image Task and Type Task (giving an overall score for total creativity) showed no significant correlations with the ETS Advanced Vocabulary test, or either reading component of the Nelson-Denny reading test (Reading Rate or Comprehension). The strength of the relationship was in all cases weak to negligible (Table 50).

Table 50: Correlations between total creativity and written comprehension measures

Correlations		TotalCreativity	ETSAVocab	ND_SSRR	ND_SSComp
TotalCreativity	Pearson Correlation	1	-.148	-.222	-.153
	Sig. (1-tailed)		.210	.111	.201
	N	32	32	32	32
ETSAVocab	Pearson Correlation	-.148	1	-.078	.376*
	Sig. (1-tailed)	.210		.336	.017
	N	32	32	32	32
ND_SSRR	Pearson Correlation	-.222	-.078	1	.054
	Sig. (1-tailed)	.111	.336		.385
	N	32	32	32	32
ND_SSComp	Pearson Correlation	-.153	.376*	.054	1
	Sig. (1-tailed)	.201	.017	.385	
	N	32	32	32	32

*. Correlation is significant at the 0.05 level (1-tailed).

7.4.6.2 Total Creativity Correlated With Written Comprehension When Controlling for General Intelligence

After accounting for the influence of general ability, the situation remained similar to that described in Table 50. Overall score for total creativity showed no significant correlations with the ETS Advanced Vocabulary test, or either reading component of the Nelson-Denny reading test (Reading Rate or Comprehension) when controlling for general intelligence (Table 51).

Table 51: Correlations controlled for general intelligence

Correlations			TotalCreativity	ETSAVocab	ND_SSRR	ND_SSComp
Control Variables	KBIT_CompSS	TotalCreativity	1.000	-.193	-.223	-.172
		Correlation				
		Significance (1-tailed)		.149	.114	.178
	ETSAVocab	df	0	29	29	29
		Correlation	-.193	1.000	-.032	.106
		Significance (1-tailed)	.149		.433	.285
	ND_SSRR	df	29	0	29	29
		Correlation	-.223	-.032	1.000	.105
		Significance (1-tailed)	.114	.433		.287
	ND_SSComp	df	29	29	0	29
		Correlation	-.172	.106	.105	1.000
		Significance (1-tailed)	.178	.285	.287	
			29	29	29	0

7.4.6.3 Type Creativity Correlated With Written Comprehension Scores

The Type Task showed a significant negative correlation (Table 52) between creativity and the ETS Advanced Vocabulary Test ($p < 0.05$, $r = -0.31$, one-tailed). However, after accounting for the influence of general intelligence, this was no longer considered significant at the 0.05 level (Table 53). In all cases, the scatterplots did not show signs of a non-linear correlation to be present. In addition, as detailed in the section on the normality of distribution, while the skew for the ETS Advanced Vocabulary was with acceptable levels, the scatterplot suggested the line of best fit was likely skewed by a low scoring outlier on the Advance Vocabulary test (Figure 35).

Table 52: Correlations between Type Task and written comprehension measures

		Correlations			
		Type_Creativity	ETSAVocab	ND_SSRR	ND_SSComp
Type_Creativity	Pearson Correlation	1	-.308*	-.175	-.281
	Sig. (1-tailed)		.043	.169	.059
	N	32	32	32	32
ETSAVocab	Pearson Correlation	-.308*	1	-.078	.376*
	Sig. (1-tailed)	.043		.336	.017
	N	32	32	32	32
ND_SSRR	Pearson Correlation	-.175	-.078	1	.054
	Sig. (1-tailed)	.169	.336		.385
	N	32	32	32	32
ND_SSComp	Pearson Correlation	-.281	.376*	.054	1
	Sig. (1-tailed)	.059	.017	.385	
	N	32	32	32	32

*. Correlation is significant at the 0.05 level (1-tailed).

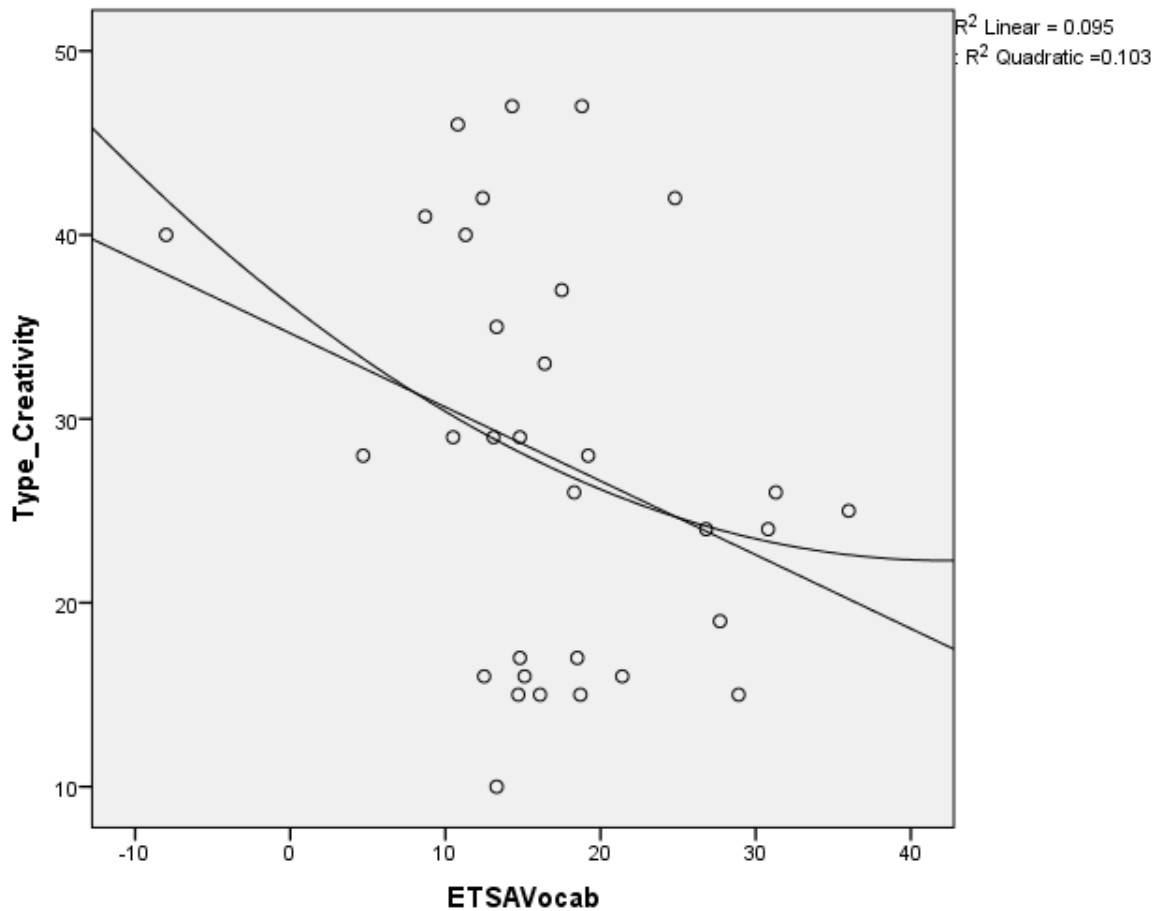


Figure 35: Scatter plot of Type Task in relation to ETS Advance Vocabulary Test

7.4.6.4 Type Creativity Correlated With Written Comprehension Scores When Controlling for General Intelligence

As described above, the significant negative correlation (Table 52) between creativity and the ETS Advanced Vocabulary Test ($p < 0.05$, $r = -0.31$, one-tailed) on the Type Task was no longer considered significant (Table 53) when controlling for KBIT Composite IQ scores.

Table 53: Correlations controlled for general intelligence

Correlations			Type_Creativity	ETSAVocab	ND_SSRR	ND_SSComp
Control Variables	Type_Creativity	Correlation	1.000	-.314	-.186	-.262
		Significance (1-tailed)	.	.042	.159	.078
		df	0	29	29	29
	ETSAVocab	Correlation	-.314	1.000	-.032	.106
		Significance (1-tailed)	.042	.	.433	.285
		df	29	0	29	29
	ND_SSRR	Correlation	-.186	-.032	1.000	.105
		Significance (1-tailed)	.159	.433	.	.287
		df	29	29	0	29
	ND_SSComp	Correlation	-.262	.106	.105	1.000
		Significance (1-tailed)	.078	.285	.287	.
		df	29	29	29	0

7.4.6.5 Image Creativity Correlated With Written Comprehension

The results found that the combined rating of the Image Task showed no significant correlations (Table 54) with the ETS Advanced Vocabulary test, or either reading component of the Nelson-Denny reading test (Reading Rate or Comprehension). The strength of the relationship was in all cases weak to negligible.

Table 54: Correlations between Image Task and written comprehension measures

Correlations		Image_Creativity	ETSAVocab	ND_SSRR	ND_SSComp
Image_Creativity	Pearson Correlation	1	.106	-.188	.064
	Sig. (1-tailed)		.282	.152	.365
	N	32	32	32	32
ETSAVocab	Pearson Correlation	.106	1	-.078	.376*
	Sig. (1-tailed)	.282		.336	.017
	N	32	32	32	32
ND_SSRR	Pearson Correlation	-.188	-.078	1	.054
	Sig. (1-tailed)	.152	.336		.385
	N	32	32	32	32
ND_SSComp	Pearson Correlation	.064	.376*	.054	1
	Sig. (1-tailed)	.365	.017	.385	
	N	32	32	32	32

*. Correlation is significant at the 0.05 level (1-tailed).

7.4.6.6 Image Creativity Correlated With Written Comprehension When Controlling for General Intelligence

After accounting for the influence of general intelligence (Table 55), the situation remained similar to that described in Table 54.

Table 55: Correlations controlled for general intelligence

Correlations						
Control Variables			Image_Creativity	ETSAVocab	ND_SSRR	ND_SSComp
KBIT_CompSS	Image_Creativity	Correlation	1.000	.029	-.179	.004
		Significance (1-tailed)	.	.438	.168	.491
		df	0	29	29	29
	ETSAVocab	Correlation	.029	1.000	-.032	.106
		Significance (1-tailed)	.438	.	.433	.285
		df	29	0	29	29
	ND_SSRR	Correlation	-.179	-.032	1.000	.105
		Significance (1-tailed)	.168	.433	.	.287
		df	29	29	0	29
	ND_SSComp	Correlation	.004	.106	.105	1.000
		Significance (1-tailed)	.491	.285	.287	.
		df	29	29	29	0

7.4.7 Correlation Between Creativity and General Intelligence

No significant correlation was found between general intelligence and graphic design creativity. Correlations were computed for each component of the KBIT-2 (Verbal, non-verbal and a composite score) for Total Creativity (Table 56) and separately for both the Type Task (Table 57) and the Image Task (Table 58). In each case, scatterplots did not show signs of a non-linear correlation to be present.

Table 56: Total creativity correlations for KBIT

Correlations		TotalCreativity	KBIT_VSS	KBIT_NVSS	KBIT_CompSS
TotalCreativity	Pearson Correlation	1	.057	-.059	-.002
	Sig. (1-tailed)		.377	.374	.495
	N	32	32	32	32
KBIT_VSS	Pearson Correlation	.057	1	.474**	.810**
	Sig. (1-tailed)	.377		.003	.000
	N	32	32	32	32
KBIT_NVSS	Pearson Correlation	-.059	.474**	1	.898**
	Sig. (1-tailed)	.374	.003		.000
	N	32	32	32	32
KBIT_CompSS	Pearson Correlation	-.002	.810**	.898**	1
	Sig. (1-tailed)	.495	.000	.000	
	N	32	32	32	32

** . Correlation is significant at the 0.01 level (1-tailed).

Table 57: Type Task correlations for KBIT

Correlations		Type_Creativity	KBIT_VSS	KBIT_NVSS	KBIT_CompSS
Type_Creativity	Pearson Correlation	1	.055	-.215	-.110
	Sig. (1-tailed)		.382	.119	.274
	N	32	32	32	32
KBIT_VSS	Pearson Correlation	.055	1	.474**	.810**
	Sig. (1-tailed)	.382		.003	.000
	N	32	32	32	32
KBIT_NVSS	Pearson Correlation	-.215	.474**	1	.898**
	Sig. (1-tailed)	.119	.003		.000
	N	32	32	32	32
KBIT_CompSS	Pearson Correlation	-.110	.810**	.898**	1
	Sig. (1-tailed)	.274	.000	.000	
	N	32	32	32	32

** . Correlation is significant at the 0.01 level (1-tailed).

Table 58: Image Task correlation for KBIT

Correlations		Image_Creativity	KBIT_VSS	KBIT_NVSS	KBIT_CompSS
Image_Creativity	Pearson Correlation	1	.036	.152	.128
	Sig. (1-tailed)		.422	.203	.242
	N	32	32	32	32
KBIT_VSS	Pearson Correlation	.036	1	.474**	.810**
	Sig. (1-tailed)	.422		.003	.000
	N	32	32	32	32
KBIT_NVSS	Pearson Correlation	.152	.474**	1	.898**
	Sig. (1-tailed)	.203	.003		.000
	N	32	32	32	32
KBIT_CompSS	Pearson Correlation	.128	.810**	.898**	1
	Sig. (1-tailed)	.242	.000	.000	
	N	32	32	32	32

** . Correlation is significant at the 0.01 level (1-tailed).

7.5 Discussion

With the dependent variable for this study (creativity output in graphic design) achieving suitable levels of reliability, it was possible to consider broader questions about the relationship between skills, creativity and domain. Namely, the main research question for this study was:

To what extent are abilities, like visualisation and written comprehension, correlated with type and image graphic design creativity, when discounting for verbal and visual IQ?

As discussed at the start of this thesis, previous work has identified that linguistic intelligence may be more important to graphic design competency than visual-spatial intelligence (Mackie, 2005). Where this study substantially differs to Mackie's research is the measure of creativity, and the tasks used. Mackie, based the assessment of graphic design competency on educators' grades for a portfolio of artwork from a student population. In contrast, this study, measured two specific aspects of graphic design creativity in isolation (type and image), drawn from broad populations (ranging from the

general public to professional graphic designers), and was only rated by judges who were currently employed full-time as professional graphic designers. It also considered the measurement of general ability, as measures of written comprehension could be influenced by this factor.

7.5.1 Graphic Design Creativity and Visualisation

Premised on the previous discussion of Mackie's study, his findings of no correlation between visualisation and creativity are fascinating. The implications of such a result are somewhat counterintuitive to many in art and design. Indeed Mackie himself suggests this was the case.

He is not alone in such a view. In design studies, Pektas (2010) considered her own findings to support Oxman's (2002) theory that "...creative emergence in design was not accidental, but depended on recognition of visual shapes and images which she called "thinking with images."" (Pektas, 2010, p.73). Furthermore, that "individuals who tended to "think with images" rather than to "think with words" were likely to be more creative in the design process" (Pektas, 2010, p.73).

One of the criticisms of Mackie's research design rested with the measures used for visualisations. In recent years, researchers' understanding of cognitive abilities like mental visualisation has become more nuanced in relation to artist and designers. For example, Kozhevnikov's et al. (2013) highlighted spatial visualisation tests are biased towards more scientific and engineering backgrounds than design and artistic professions. In Study 5 of this thesis, this was addressed by the inclusion of the Snowy Pictures test and Mark's VVIQ, alongside a traditional form of visualisation through the Purdue PSVT:R.

Despite these considerations for research design, no significant correlation was found between visualisation and graphic design creativity. The results of this study found that the combined rating (total creativity score) of the Image Task and Type Task showed no significant correlations with the Marks VVIQ, Snowy Pictures, or the PSVT:R. The strength of the relationship was in all cases weak to negligible, and after accounting for the influence of general intelligence, the situation remained the same. Equally, taken in isolation, no significance occurred for both the Image Task and the Type Task. In all cases, it was considered that the scatterplots did not show signs of a non-linear correlation to be present.

Given the findings above it would appear that despite two very different approaches to research design, both Mackie's original findings and the new results described here, there is no correlation between graphic design creativity and visualisation.

7.5.2 Graphic Design Creativity and Written Comprehension

For the most part, the results of Study 5 found that the combined rating of the Image Task and Type Task (giving an overall score for total creativity) showed no significant correlations with the ETS Advanced Vocabulary test, or either reading component of the Nelson-Denny reading test (Reading Rate or Comprehension). With one exception, this result occurred for both the Image Task and the Type Task. The Type Task showed a significant negative correlation between creativity and the ETS Advanced Vocabulary Test ($p < 0.05$, $r = -0.31$, one-tailed).

In broad terms, this was the complete opposite finding to Mackie positive correlation between linguistic intelligence and graphic design competency (of which creativity was a component). What could not be accounted for in Mackie's research design was the

influence of general intelligence on his results.

As CHC Theory highlights, test scores for written comprehension and writing ability (the kind of abilities tested through Mackie's operational definition of linguistic intelligence) are correlated with general intelligence scores. It was possible that Mackie's results were also an indicator of a correlation between academic grade for graphic design work and general intelligence. With Study 5, however, after accounting for the influence of general intelligence, there was no longer a significant correlation between the Advance Vocabulary Test and the Type Task creativity score. Indeed, the scatterplot suggested the line of best fit was likely skewed as a negative correlation by a low scoring outlier on the Advance Vocabulary Test.

Given Study 5 findings as a whole, this was the only instance when the use of the KBIT-2 general intelligence score was able to clarify the results. The value of this for the research design, however, should be considered in relation to the practical impact of undertaking the fieldwork for Study 5. For example, the addition of the KBIT-2 introduced a level of additional financial costs into the study, and increased the time required by participants. Both factors were influential on the number of participants that could be recruited to this study: practically, in terms of budget constraints, but perhaps more importantly, though less directly, with regard to the working lives of professional graphic designers; the additional increase in time placed a barrier to participation, and in practice took longer to administer than was suggested.

It is worth stating that, in order to gain access to KBIT-2 from Pearson, a researcher is bound by a contractual obligation to maintain the confidentiality of test administration details, scoring, etc.). Without negating this obligation, it appears to be publicly available

knowledge, both via books and the web, that this test follows a discontinuation rule. The reason to highlight this here is that the more answers a participant gets correct within the discontinuation rule, the longer the test takes. Of the participants in Study 5, the large proportion would be considered to have average or above average IQ scores (that is, a standardised score above 100). In hindsight, perhaps this was predictable. Whilst data on a participant's highest level of academic qualification was not recorded or required for this study, given the recruitment of professional designers, students undertaking a graphic design degree, and students who may have taken, or are currently taking degree level study in psychology, it is likely that the majority of participants would fall within, or above, an average IQ score, and testing times would increase as a result.

As the KBIT2 is not a timed test (though procedures are in place to offer prompts, to keep the pace within broadly expected timings -that is under 30 minutes- and take into account levels of stress); it is also a test that was likely to be unfamiliar to professional graphic designers and graphic design students (less so in this sample student from psychology, recruited via the OU). For example, not discounting the influence of an individual's speeds of decision making, many of the participants with a design background found the test highly absorbing, particularly the visual matrix section, and riddles. As required, to keep the pace, prompts were offered to make guesses, but equally, replies were given that they did not want to guess, or they wanted to take their time (given the non-timed nature of the test, such replies were respected). Indeed, when dealing with a confident design professional, especially creative directors or other senior level positions with design agencies, it is likely that the personality traits of such individuals make them strongly self-determined when faced with a novel challenge. The impact of which is that the KBIT-2 took longer than expected, in some cases up to 45 minutes, much greater than the administration manual predicts. Added onto the other psychometric measures, and whilst

this was the last task participant undertook, fatigue was noticeable in some instances; particularly at the end of an intensive day of professional practice. This suggested limitations for Study 5, and consideration for the research design in future studies.

7.5.3 A Graphic Design CAT Without Pre-Selection of Artwork

One of the subordinate research questions for study 5 was: Can the Type Task, and Image Task, remain reliable measures of graphic design creativity without pre-selection of artwork? As highlighted in Study 3 and 4, with specific tasks, the CAT could be a reliable measure of graphic design creativity, but a number of caveats were required to achieve this, one of which was the creation of optimal conditions by sampling artwork prior to CAT assessment.

In Study 5, without pre-selection of artwork, Cronbach alpha for both the Type and Image Task creativity remained above 0.7. Indeed, each was considerably more: for the Image Task it was 0.81, and for the Type Task it was 0.92. In both cases, these were higher levels of consensus than in Study 4, which ran exactly the same tasks, had exactly the same number of judges, and virtually identical instructions.

Some adjustments of CAT protocol were made in Study 5 relative to Study 4, notably, the verbal re-emphasis to judges not to compound their rating of creativity with technical execution or aesthetic appeal. If this did have an influence then it is more the case for the Type Task (with the alpha increasing from 0.82, in Study 4, to .92 in Study 5), however, this comparison needs to be considered alongside the fact that judges were different in Study 4 and 5, as was the artwork being judged.

A further point was that the population of participants in Study 5 were more diverse

(ranging from members of the general public with no art, design or craft experience, through to full-time professional graphic designers with decades of commercial practice), and this may have impacted on the higher level of consensus. Again, the influence is more directly inferred from the Type Task rather than the Image Task, with alpha levels that have stayed much the same as Study 4.

Notwithstanding the limitations of a direct comparison between the Type Task in Study 4 and Study 5, the increase in consensus from 0.73 to 0.92 is likely to be statistically significant. Perhaps, if this is the case, then it lends further support towards the benefit of the "creativity only" instructions, the value of verbal re-emphasis, and diversity of artwork/population. These caveats to CAT protocol for graphic design creativity do not seem detrimental to research design, and would appear to improve the reliability, if not the validity, of the CAT as a measure of graphic design creativity.

7.5.4 A Combined Creativity Score for Type and Image

A second subordinate question of Study 5 was: To what extent are the Type Task and Image Task correlated with each other? Given previous results, the expectation was that a relationship would exist between creativity on the Type Task and the Image Task, but that this would be unlikely to be strong. Given the findings of Study 5, the correlation was significant ($p < .05$), as it was in Study 4, and showed a moderate positive relationship ($r = 0.32$).

The fact that in both Study 4 and 5 a positive relationship between the Type and Image Task occurred does give justification for their inclusion, and use in design education and assessment. However, it will be of interest to some researchers and educators that the

relationship is not a strong one. Clearly, this is only two studies, but the average would suggest this relationship is a moderate one. A moderate correlation could partly explain the challenges experienced by previous CAT studies of graphic design, those that set tasks that integrated image and type within one task (i.e. Study 2, the T-shirt Task, and Wojtczuk's (2014) eco-poster task.). In practice, while both image and type form considerable visual elements of professional graphic design, and graphic design education, from a creativity design research perspective as yet it is not clear what design brief is suitable for the CAT that integrates these both type and image.

From a research design perspective, a moderate positive relationship between image and type could be seen as an asset relative to a much stronger correlation. It would appear each task is measuring a different facet of graphic design creativity, and that a high score on one does not mean a high score on both. Viewing the scatterplot (Figure 36) of the Type Task creativity scores correlated with the Image Task creativity scores it is clear that a middle portion of results shows participants who excelled on the Type Task with average results on the Image Task.

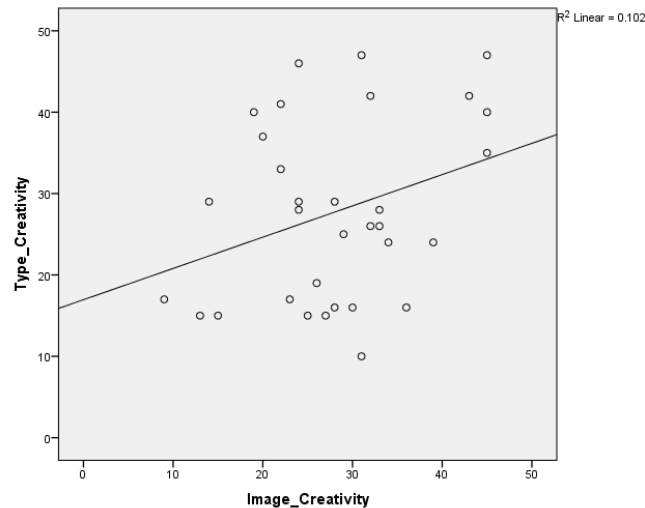


Figure 36: Scatter plot of Type Task to Image Task creativity scores

7.5.5 Correlations Between Creativity, Technical Execution and Aesthetic Appeal

As detailed in the method section of Study 5, judges were asked first to assess creativity only, and then randomly assigned to judge technical execution or aesthetic appeal. A final assessment was either technical execution or aesthetic appeal depending on the previous outcome. A final subordinate question was: To what extent are technical execution, aesthetic appeal correlated with graphic design creativity?

For some time, guidance has been that when a CAT task is developed for a different domain (i.e., one that has not been studied with a particular task), researchers should ask judges to rate both technical execution and aesthetic appeal (Amabile, 1982; Hennessey, 1994), and check to see creativity ratings are distinct from these criteria. Once this has been shown to be distinct, researchers need only ask for ratings of creativity, and can assume that technical execution is no longer a consideration for this task.

Indeed, in Amabile's 1982 work the extent to which creativity may be isolated from such factors was a formative part of her paper, and she concluded that "...although judges were not provided with a definition of creativity...they consistently and reliably identified a quality

in both types of product that was distinct from technical execution. Moreover, for artworks, it was distinct from aesthetic appeal as well" (p.1010).

At the same time, Amabile acknowledged that for some domains the distinction between technical execution and aesthetic appeal may be less clear and that creativity is likely to correlate with these aspects of the work. This would certainly appear to be the case in this example of graphic design creativity.

For the Type Task, each correlation was positive at a significance level of 0.01 (1 tailed), and each suggestive of very strong correlations: creativity and aesthetic appeal ($r = 0.93$); aesthetic appeal and technical execution ($r = 0.90$); creativity and technical execution ($r = 0.87$).

For the Image Task, each correlation was positive at a significant level of 0.01 (1 tailed), with the strongest correlation between aesthetic appeal and technical execution ($r = 0.82$), followed by creativity and technical execution ($r = 0.68$), and creativity and aesthetic appeal ($r = 0.48$).

It would appear that the Type Tasks, and to a lesser extent the Image Task, are clear examples of situations where creativity, technical execution and aesthetic appeal are highly correlated. Furthermore, evidence can be found that a substantial single component likely to be present that underpins the judges' assessment of creativity, technical execution, and aesthetic appeal.

The results of the factor analysis highlighted that, for the Type Task, only one principal component with an eigenvalue greater than one was identified, with a single component

that could explain 93% of the variance in the data. Additionally, for the Image Task, only one principal component with an eigenvalue greater than one was identified, and this single component explained 78% of the variance between creativity, technical execution, and aesthetic appeal.

7.5.6 Emergent Considerations

One issue that emerged from the analysis of Study 5 was the level of detail for a graphic design task/brief. For example, the creativity tasks in Study 5 did not clarify which format the A4 sheet should be presented in; the majority of artworks from participants were landscape, whilst a few others chose portrait. To what extent such variation in format, or in some cases the layout of the graphic on the sheet, may influence the judges' consensus on creativity appears not to have had a direct impact overall on consensus, but it is likely that such choices could have had an influence on views of technical execution and/or aesthetics appeal.

From a design perspective, such choices by participants to opt for landscape or portrait was not problematic. The challenge came from the administration of the CAT. In the creation of a digital file to show judges, the artwork in portrait format was compounded by the default setting and limitations of PowerPoint. These constraints had not been apparent in the previous studies. Unfortunately, it was not straightforward to switch between landscape and portrait in a single PowerPoint file. While this can be achieved in presentation mode by switching between two documents that are linked (one with portrait files, the other with landscape files) to create the illusion of a single file, this did not work when converted to a single PDF file. As detailed in Study 4, PowerPoint was a means to randomised artwork in a single PDF file, but Study 5 highlighted that the software was unable to do so without all pages conforming to either all standard portrait, or all landscape format, but not both.

For this reason, Adobe Illustrator was used instead. However, having created a PDF with both landscape and portrait artwork shown correctly, the challenge of randomising the artwork in Adobe Illustrator was more difficult. Unfortunately, the process did not adjust for the differences in page orientation and created a file only in landscape (despite the selection of the auto portrait/landscape option). For this reason, the artboards for each artwork in Adobe Illustrator needed to be reordered by hand for each judge. This process appears not to be amenable to automation, and whilst laborious did achieve the PDF file showing a random order of artwork, and in the appropriate page orientation. Whether such additional procedures are worth the effort of offering participants the opportunity to select portrait and landscape formats is something to consider for future studies and discussion. For graphic design creativity the restriction to, for example, landscape only formats, for future research may be considered too restrictive a condition within the art and design community.

A final point on the issue of portrait or landscape format came from observing the judges as they undertook their assessments. Judges had four elements to engage with: an A4 hard copies of the design brief and instruction for rating the work; a laptop showing a PDF file that had all the artwork shown in the correct format (either landscape or portrait); an A3 laminated rating sheet; a set of laminated thumbnails of the artworks shown on the laptop that they could move around on the A3 rating sheet. What was noticeable, was that some artwork shown as portrait would occasionally find themselves in a landscape position on the rating sheet. Equally, as the rating progressed from creativity to technical execution, to aesthetic appeal, for example, judges tended to spend less time looking at the PDF file on the laptop and more time concentrating on the rating sheet. It was for this reason that each time after a judge had first assessed the creativity of the artworks, they

were reminded that the details, both technical and aesthetic, would be better found on screen than in the thumbnails. It is perhaps a minor point given the high levels of consensus achieved in this study, but some artworks that were portrait may have suffered from being viewed as landscape thumbnails, and judged more harshly. Especially, in those instances where the portrait format was crucial to understanding the work, it is difficult to rule out they may not have been disadvantaged by the paper based rating sheet, and the dominance of predominantly portrait format artworks. Whether this is enough of a rationale to standardise the type and Image Task to a landscape only format, is difficult to say.

This issue may appear overly detailed, but it reflects a broader point regarding how standardized should graphic design creativity tasks be, or could become: did the current instructions for these tasks give an opportunity for creativity because they did not specify a standard format to be followed by all participants, or does the lack of a standard format introduce more complexity and distract judges unnecessarily? It was possible to standardise all the artworks judges would view through opting for a landscape format. However, whilst this may enable easier administration and allow judges to view more easily all the artwork without the distraction of changes in format, for some artworks the choice of landscape or portrait was part of the quality of the work.

7.6 Conclusion

Study 5 was built upon the research of Cattell, Horn & Carroll's (CHC) theory of human intelligence, and the Consensual Assessment Technique (CAT) in relation to graphic design creativity. This was the first time both these significant models and methods had been applied to graphic design.

The outcome was to explore to what extent abilities, like visualisation and written comprehension, correlated with graphic design creativity, when discounting for general intelligence. Where previous studies had found correlations, in this study, no significant correlations occurred when general intelligence was accounted for.

At the start of this thesis, one of the underlying assumptions regarding the relationship between competencies (be they defined as skills, abilities or knowledge) and creativity was that there was a relationship. Several authors reflect views on the nature of this relationship that are broad and diverse: whether these are relatively simple positive or negative linear correlations, or a more complex form of non-linear relationships; the assumptions is one exists. The results of this research, suggest such underlying assumptions may not be correct when tested against empirical study.

If this is true, this could have a direct impact on creativity training and educational practice. For example, what the results suggest is while there is not a correlation within the group as a whole, there is in specific cases. For three participants, scoring highly on the type graphic design creativity task was matched with scoring highly on Marks VVIQ; conversely, for two higher creativity scoring participants, this was the complete opposite with lower VVIQ scores. Such a pattern was repeated for lower creativity scores and those participants with average creativity scores. As cannot be emphasised enough, the adage that correlation is not causation is important to stress; improvement in visualisation ability is not necessarily going to improve graphic design creativity, for some it may, for other not. What such results highlight is that fastidious views on what abilities are important to nurture can be a theoretical stance not fully borne out by empirical study, but based on isolated cases that can shape individual and institutional perceptions.

Chapter 8: General Discussion and Conclusion

8.1 Introduction

In previous discussion sections of the thesis, the discussion has been self-contained within a particular study, either Study 1, 2, 3, 4 or 5. For the most part, this is because each study addressed specific challenges and details that required resolution in order to consider to what extent specific cognitive abilities may underpin creativity in graphic design.

This chapter broadens the discussion beyond the immediate findings of any single study, and will explore larger arguments and insights that may be present; such as, how findings may compare across all studies undertaken for this thesis, the interaction between other scholars findings, and broader observations/future research opportunities that may be inferred from this analysis. In this regard, this discussion comments of the thesis as a whole. It also returns to the three larger themes of Theory, Education and Industry (first highlighted as the start of this thesis), and reflects how the findings of this thesis may have a broader impact.

8.2 The Art and Science of Creativity Research

Many academics and scholars have dedicated their careers to the study of creativity. For those outside of this section of academia the general response to the study of creativity is amazement that such a field exists. What is even more remarkable is the research that has inspired this thesis can be traced back to the early 1950's. While creativity research may be new to some, the topic is far from new to critical debate and discussion. Indeed, international publications like the Creativity Research Journal, Journal of Creative Behaviour, the Psychology of Aesthetics, Creativity and the Arts, have been specifically

focused on creativity studies for decades. Within the discipline of design academic publications like *Design Studies*, *Design Research Journal*, *Co-Design*, *Design Journal* have each had an ongoing interest in creativity for many years.

And yet, despite this, advocates of creativity research regularly encounter what can be described as an ongoing negative bias towards the study of creativity. Particularly, when the topic is framed around the science of creativity. For some, this concept is laughable. For others, it is a dangerous development likely to undermine artistic practice. These views are understandable. What is creativity is a question that has puzzled great minds for centuries. Some believe this answer is unknowable, that creativity is an innate gift; you either have it or you do not. This thesis, however, contributes to a growing community of scholars that believe through scientific research we can understand this unique ability further. That creativity research offers profound knowledge to help enhance our potential for creative thinking, problem solving and innovation.

Confusion and negative bias are caused partly because of the assumption that what one person defines as creativity is what everyone else defines as creativity. It often is not the case. This is not to imply creativity is undefinable, but to highlight that for an informed public discussion more refined use of the term creativity is required. Being explicit on definition is challenging, but, creativity research has much to offer, with fundamental ideas and theories that are useful.

One model to highlight (and discussed at the start of this thesis) is that creativity research can broadly be categorised into four distinct areas of study: these have become known as the 4P's of creativity research (Isaksen & Murdock, 1993). They are: The Person, The Process, The Product and The Press (the creative climate, culture or environment). The

key point is there are different views on how to define what creativity is. The issue, then, is that depending on viewpoint, debate may disagree or agree not on quality and merit of a research study but because the same or different perspectives are held.

The stance taken in this thesis is to define creativity through the perspective of a creative output (product) and to assess this creative output through use of the Consensual Assessment Technique. Many creativity researchers consider the creative product to be a useful construct when it comes to researching creativity. The reasoning is that in order to validate whether someone has a creative process, personality or been influenced by a creative environment, then the validation of the creative product will likely occur at some point (Kaufman & Baer, 2002).

A further theory to consider is the 4C's model of creativity. The work of James C. Kaufman, and Ronald Beghetto, highlights that it is useful to see creativity in four different but interrelated ways: mini-c creativity, little c creativity, pro-c creativity, and big-c creativity. For them, mini-c is defined as "the novel and personally meaningful interpretation of experiences, actions, and events" (Beghetto & Kaufman, 2007, p.73). It is not miniature creativity in the sense of importance but in the amount of social recognition the "creative output" receives. After mini-c, little-c, pro c and big c are each characterised by differing levels of social recognition. The achievements of great minds that have fundamentally shaped the world we live in would be associated with big-c creativity. Pro-c is recognised within a profession, and little-c is associated with educational achievements.

It is possible to be disturbed by this model as it could infer that anything less than big-c creativity is lesser, or watered down creativity, and that emphasis on social recognition can be a point of controversy. Kaufman and Beghetto are likely to advise not to see the

four-c model in that way, that the value of these distinctions is to acknowledge how levels of social recognition impact on the criteria for creativity and its assessment. Indeed, it is important to consider this spectrum when debating creativity, and not conflate these. This thesis has mostly focused on the products (creative outputs) of pro-c creativity.

In its richness, and relative youth, creativity research offers profound knowledge to help enhance our potential for creative thinking, problem solving and innovation. Moreover, if the future develops into the automated, Artificial Intelligent driven knowledge economy some think it will, then the ability to be creative will be a crucial asset, perhaps “the” crucial asset. Indeed, at the first UK creativity researchers’ conference in Edinburgh (2017), in his keynote speech Prof. Giovanni Corazza went as far as to say that future generations will look back, and “in retrospect, *Creativity* will be recognized to have been the most important field in *philosophy, psychology, and cognitive sciences* for the XXI century”. That is a major claim and if he is correct then the science and study of creativity will become a serious business.

8.3 CHC Theory and Graphic Design Competencies

From the perspective of government, higher education and industry the transition from design higher education into professional creative practice is an ongoing debate. The discussion has tended to centre on graduate quality and one view frequently expressed is that the graduates produced are not ‘good enough’ for the tasks employers require of them. For example, a 2011 Skillset report on this problem highlighted that in the Creative and Cultural Industries: “*Employers in the industry state that the higher education system is not fit for purpose; supposedly highly-qualified graduates are not entering the sector ‘job-ready’, and thus need to acquire further skills*” (p. 38). This disconnect between higher education and job competence has been described as a ‘skills gap’, and while there is

evidence of skills gaps in different industries, Jackson (2016) notes, there is "a limited number of relevant studies" (p. 200). As a result, the evidence base for skills gaps in creative industries research has tended to favour survey methods, and there remain difficulties in the empirical study of skills gaps in terms of what measures to use.

In chapter 2 of this thesis, it was highlighted that O*NET's taxonomy regarding competency, abilities, knowledge, skills and tasks helps clarify the conflated usage of the term skills. Indeed, Levy and Hopkins (2010) example of high-level skills as "...a broad concept reflecting an individual's ability to use tacit knowledge to assimilate and interpret information" (p.32) appears confused from the vantage of O*NET, F-JAS and CHC Theory.

By its size and scale, the focus of this thesis on visualisation and written comprehension in Study 5 and its relationship to graphic design creativity is a small contribution to the skills gaps debate in the UK. It is, however, substantially different to a survey approach or qualitative research on employers, employees, educators or graduates perceptions of skills gaps. In this respect, how the study was designed (its operational definitions, theoretical underpinnings, and sampling) is as interesting as its results and findings.

8.4 Growth of CAT Update

Since the publications of Study 1's findings in 2012, the CAT has continued to follow an upward trend in growth within scholarly journals. Indeed, more CAT studies and citations have occurred in the five-year period between 2010-2015 than in the previous 28 years of CAT research (Figure 37). At continued levels of growth, a conservative estimate is that, by the end of 2016, over 2000 CAT studies and citations will have taken place since 1982, of which 63% will have occurred within the past five years. In this respect, Study 1 (whilst

still remaining the first and only systematic review of the CAT in design, and based on 722 studies prior to 2010) is already considerably out of date. This is all the more reason that the findings of this study are valuable to a community of scholars both with and outside of design.

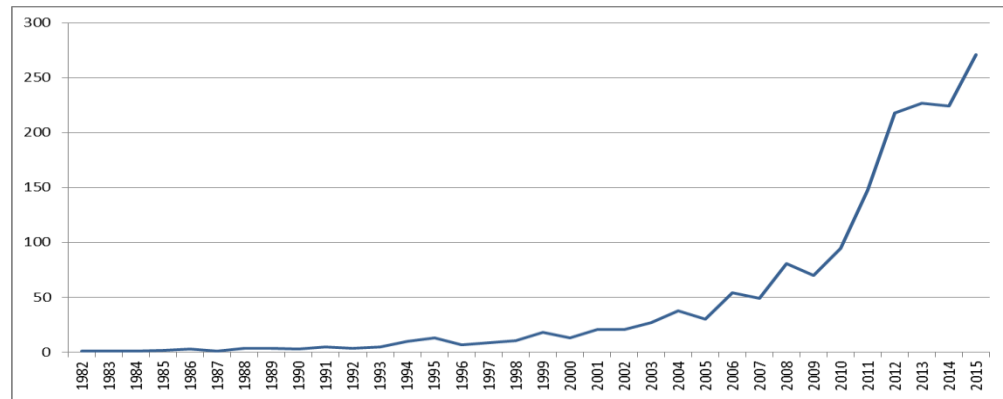


Figure 37: Number of CAT citations updated to 2015

It is important to highlight that the dependent variable, and key measurement through this thesis was creativity, and specifically creative output. Without the robustness of how this construct was measured, the findings for this research would be brought into doubt. Equally, this is an argument to apply to the validity of previous studies and findings in relation to other scholars of design and creativity research. Regardless of the statistical power, significance and correlation, if the operational definition and protocols are in doubt (especially in a study where creativity is the dependent measure), then findings are open to greater criticism.

It is for this reason that the level of attention to procedure and precision of details regarding the CAT's application to graphic design creativity were present in this thesis. For some, this degree of detail may be laboured (especially given such discussion is not always readily present in other scholars application of the technique), but for this thesis, it can be argued that such details matter. Whether this was needed because the

assessment of graphic design was particularly sensitive to such nuances is possible, and it may be that in other domains this degree of detail is not required. However, as is the case in other areas of design, like ergonomics, the value of designing for limiting users of a given product, offers a level of design consideration to create a better "product" for the 95 percentile who do not fit within the extremes of a population, as well as those outside of it.

8.5 Novelty of CAT to Design Studies

As has been highlighted in the systematic literature review of Study 1, at the time of this study few published works existed on the CAT as a measurement of design creativity, and even less relevant to graphic design creativity. Of those graphic design studies that existed, levels of consensus were inconsistent, and in some studies were considerably below the threshold for reliability; an issue that was present in Study 2 for this thesis was that the CAT did not seem to work, not consistently.

Of those researchers using the CAT in graphic design, prior to the 2012 International Conference on Design Creativity (ICDC), scholars were not aware of each other's works: finding had yet to be published, and full details were not readily available until 2014 or at the following 2015 ICDC conference. Once this data was available, it was clear research findings were highlighting a graphic design CAT appeared more sensitive to methodological protocols than previous studies in other domains had reported. A consideration is that had this data not been the result of PhD studies, then the reporting of these findings many never have been made publicly available, and assigned to the proverbial "bottom draw". However, access to these results served as an encouragement to understand further what may be influencing these low levels of consensus and to revisit past assumptions upon which the CAT was established. Two of these assumptions are

worth further discussion and are interrelated. The first is the evidence for “creativity only” instructions, the second is the level of correlation between creativity, technical execution and aesthetic appeal

8.6 Creativity Only Instruction

Returning to the findings of study 4 and 5, a distinction between the CAT protocol of Study 4 with that of Study 5 was the verbal re-emphasis of creativity only instruction to judges. The previous debate at the ICDC2015 reflected whether the judges had read, sufficiently, the written instructions given in Study 4, and whether the wording of the instruction was strong enough to get the point across about creativity only assessment. By verbally re-emphasizing this aspect of the instructions, Study 5 can be argued to have addressed both concerns for awareness and emphasis.

From a research perspective, study 4 and 5, offer three different approaches to CAT instructions. The first set (the adapted Kaufman's et al instruction) gave no instruction other than to assess the creativity of the artworks. The second set (the adapted Baer's instructions) directly addressed this issue of creativity only instruction but did not verbally re-emphasise these. The third set did verbally re-emphasise the creativity only instruction.

Specifically, the results for the Type Task of Study 4, were judges who received the adapted Kaufman et al. instructions had an alpha of 0.73; judges that received the adapted Baer's instruction (to discount technical execution), an alpha of 0.82; and study 5, (with verbal re-emphasis) an alpha of 0.92.

For the Image Task, the adapted Kaufman's et al. instructions had an alpha of 0.75; for the adapted Baer's instruction the alpha was 0.80; for the verbal re-emphasis, 0.81. Table

59 below, helps clarify these results for comparison.

Table 59: Increase in alpha across study 4 and 5

	Study 4 (creativity)	Study 4 (creativity only)	Study 5 (creativity only)
Type	0.73	0.82	0.92
Image	0.75	0.80	0.81

8.7 Correlations Between Creativity, Technical Executions and Aesthetic Appeal

Given the high level of correlation, in Study 5, between creativity, technical execution and aesthetic appeal, what does it actually mean that they are highly correlated? And, what happens in those domains where they are correlated: should researchers not use the CAT?

On one level it may be possible for detractors of the CAT to see this evidence of confluence as a basis to say the method does not measure creativity. This is not the position of this thesis: the argument here is that assessing creativity output is complex (especially graphic design creativity, and likely for other areas of design).

Indeed, it is that complexity that meant the CAT when first applied in Study 2 and in Wojtczuk (2014) did not work. With the various caveats considered in this thesis, clearly, acceptable levels of consensus can be achieved, but one of those caveats may be how important it is to instruct judges not to consider technical execution and aesthetic appeal in their ratings of creativity.

The case made here is further support for why the caveat on technical executions and

aesthetic appeal is important. These constructs would appear to be highly correlated for certain domains, but this does not mean that judges cannot give ratings on creativity only. What is required is that this instruction is made clear to them.

Emphasising that caveat enables a judge to tacitly reduce the background "noise" in creativity rating that is present from technical execution and aesthetic appeal. This may be a challenge, but the evidence of Study 5 and Study 4, is that they can achieve this. In contrast, if research does not do this, then every study could be skewing the creativity aspect with technical execution and aesthetic appeal. That has implications for comparison across studies.

8.8 Distinction Between Aesthetic and Technical in Other Design Domains

A key question to consider is, do the original guidelines on the distinction between creativity, technical execution and aesthetic appeal apply to graphic design? Again, it must be acknowledged that Amabile made very clear this would not apply to some domains. Equally, the CAT was foremost used in social psychological research contexts, and showing discriminant validity between creativity and other compounding constructs was important to establish the validity of the measure at that time. Never-the-less, contradictory evidence in Amabile and Hennessey's own work raises doubts about creativity and technical execution as distinct constructs. Within Amabile's 1982 studies she found correlations as high as .77 between creativity and technical goodness, and Hennessey (1994) presented statistically significant correlations as high as .71. As identified previously, other contemporary CAT based research studies more directly related to professional design, such as Valgeirsdottir, Onarheim & Gabrielsen (2015), have also identified high positive correlations, yet other design creativity researchers make no mention of having considered this in their research design. (Lee, Gu & Ostwald 2015).

8.9 Not Meeting Assumptions

If assumptions of distinction are not met; but we have a set of creativity instructions that minimise the background "noise" of technical execution and aesthetic appeal, then is rating all these constructs really worth the amount the time and effort of future design researchers who use the CAT? For example, Valgeirsdottir, Onarheim, & Gabrielsen, (2015) asked judges to rate three other constructs and creativity; Study 5, here, asked for two other constructs. This is a considerable amount of work.

While these are concerns to be debated within the broader community of scholars, for the value of starting that debate, the conclusion of this thesis is that (dependent on the domain, the research aims of a given study, and with the caveats discussed), if design creativity researchers only want to measure creativity they should do so, and without this being viewed as a methodological flaw because of the original guidelines suggested by Amabile and Hennessey. The investment in a judge's time and effort, and the researcher's time and effort, make additional assessments unproductive. Recruitment to CAT studies is an acknowledged challenge as it currently stands, when the likelihood is there will be a highly significant and strong correlation between creativity, technical execution and aesthetic appeal, it is questionable if that additional effort is worthwhile for future design creativity studies.

8.10 Is Creativity Domain Specific?

The argument for creativity only instructions (given the high levels of correlation between creativity, aesthetic appeal and technical execution) raises the broader issue of the extent to which creativity may be domain-specific or not.

As highlighted at the start of this thesis, the domain specificity of creativity has captured the interest of many scholars for a number of years. In themselves, as important as these theoretical discussions are, they also have implications for creativity training and practical strategies for creativity enhancement.

One distinct finding has been that the measure of creativity used in a study appears to influence the outcome of asking if creativity is domain-specific or not (Plucker, 2004). For example, when studies focus on creative products evidence for the domain specific nature of creativity is found (Silvia, Kaufman & Pretz, 2009). Conversely, when studies focus on the creative person, evidence for the domain general nature of creativity is found (Plucker, 1999). Due to this, contradictory findings on the domain specific/domain general debate need to be considered in the light of methodological choice regarding how to assess creativity that differing studies have used. Without this insight, it is quite reasonable for a scholar to evidence that creativity is domain specific, or domain general, based on the choice of methods selected.

In subsequent years, one impact of this bias has been the development of compromised models, such as Amusement Park Theory (APT theory), that have integrated both the domain specific and domain general nature of creativity within single models. APT Theory (Baer & Kaufman, 2005), proposed the analogy of entering an amusement park with a wide range of activities available and, for a range of reasons, a person may be predisposed to some activities more than others. A person's aptitude for an activity (for example, throwing an object to hit a target), means that they may do well in many tasks that require this kind of ability (throwing/targeting in general). But, they may show a particular aptitude for a given task that has a certain weight and shape of object to be thrown, or distance from a target, or any other specific features.

The analogy is useful, and particularly if the amusement park is replaced with the field of creativity research; the activities replaced with psychometric tests and methods of creativity assessment. Enter the amusement park of creativity research and one exhibit is the TTCT, it has various tasks to undertake, but in essence is contained under a single stall. Other large exhibitors are Pearson Clinical, with the KBIT-2, or Educational Testing Service (ETS) Kit of Factor-Referenced Cognitive Tests.

This is not the case for the CAT, it has multiple exhibitors, for the most part each with their own specific tasks, and instructions. Due to this, a rich diversity of approaches have evolved that appear to be known, or promote themselves, under a single CAT brand. This is both a strength and weakness, but when such an environment is used for creativity testing it is quite possible to gather contradictory results, and understandably, comparison can be fraught.

The results of this study offer a set of tasks for the CAT (this time for graphic design), and a further interpretation of instructions and protocols. As discussed throughout this thesis, the application of the CAT to graphic design was not straightforward. Had it been, it is possible the questioning, testing, and depth of detail undertaken would not have occurred. It would also have been possible to continue to add to the industry of CAT research without standing back to question it.

New tasks and improvement in protocol in themselves have value to the field of design creativity assessment and creativity research more broadly. However, given the challenge faced to create a graphic design CAT in this study, it highlighted that the development of CAT research has to date been analogous to a cottage industry. Clearly, as a field of study it has been significantly shaped by researchers like Amabile, Kaufman, Baer and

their colleagues, but it has also been extended to numerous domains and uses beyond its initial intentions.

Such richness and openness during the last 30 plus years can be seen as positive, but the argument presented here is that it is now timely to fundamentally review the field of CAT research as a whole. If CAT research is to further advance itself as a robust measure of domain creativity, then a meta-analytic review is required on which to engage scholars on the standardisation of research protocol. Without such standards, scholars understanding of the science of creativity and the domain specificity or generality of creativity is likely to be undermined; specifically, the interaction between domain-general and domain-specific creativity. Creativity research has a substantial number of CAT studies, what it does not have, at present, is a study of those studies: a meta-analysis.

8.11 Ethics of Showing Creative Outputs

"...I really don't want to know how I was judged on the two creativity tasks (too scary)"
(anon)

At the end of the data analysis for Study 5, all participants were given the opportunity to see their results. In the extract above, this expert participant was very interested in the other psychometric measures and their scores but requested the creativity results had to be taken out. Indeed, for a professional graphic designer to risk finding out how their creativity was assessed by eight peers, and within the academic standards aspired to in this research study, is intimidating. Despite the fact, no one else, other than the key researcher, would know the results (and it was made very clear this information would be treated in strictest confidence) it takes a great deal of courage on the part of professional designers to take part in a study like this. They are putting themselves on the line:

anything less than scoring in the highest quadrant is likely to be annoying at best but more likely to be substantially upsetting.

The key issue to highlight here is around the confidentiality of those taking part, and why the ethical principle of not showing artworks has been at the core of Study 4 and 5. Such a stance is unusual for design creativity, and at the time, relative to other studies, perhaps seemed unnecessary.

At least two factors are clearly at play: firstly, as has been discussed in Study 1, the majority of design studies, especially those interested in creativity, have focused on the creative processes of designers and students of design. This is a generalisation, and the growth in CAT studies may be changing this situation, but historically, it is a reasonable statement to make. Thus, many design studies focused on creative process have not had to address the ethical issues of self-labelling and the impact on creative identity that a "product" focused research study must address. Secondly, given the challenges of recruiting to studies like this one, and CAT studies more generally, when creative outputs are available they offer a unique insight: often an image can say much more than can be written. Understandably, within design research and creativity research broadly those studies that do engage with creative outputs have visual findings to show, and want to celebrate this. Such a desire seems harmless enough with exemplars of excellence (who would not want the accolade of being the "most" creative work within a rigorous academic research study on creativity). The ethical issue arises with those works that do not get shown, as much as with those that get labelled as examples of "low" creativity.

Hopefully, this argument is not too abstract to accept, and the practical implications are clear enough. Some creativity researchers and journal editors will not like the implications

of this suggestion; it reflects poorly on past choices and infers a lack of considerations for participants that is likely unwarranted. However, the position taken, based on the experience of undertaking this study, is that it is unethical to show any artwork from a study like this, both on the grounds of negative self-labelling (a concern clearly stated in the BPS code of ethics) and for the future sustainability of creative output focused design research. As this field grows, and the CAT method becomes more popular within design research, it is crucial not to build a reputation for negative associations and anxieties in the minds of the very participants required for this type of research. It is in all researchers' interest to make sure this does not happen.

Likely some form of compromise position between showing no examples and full disclosure can be found in the future. Hard evidence that participants in this study took part directly because they had clear assurances their work would not be shown is not available, but the tacit knowledge left from this experience, along with the quotation given above, is that this is the case. It is implied that it would be much harder to recruit had this not being the approach taken.

Why this may be so is difficult to reflect upon. Indeed, there is something of a contradiction here, as many design professionals are regularly judged by clients and peers on their creative output, more than their process. Also, students on design courses to a varying extent, depending on the vocational focus of their course, are so familiar with the assessment of their creative outputs that this is just another "crit". It is interesting that in previous studies on creativity in games design (Jeffries, 2009, 2011) one of the competencies expressed as important to acquire for creativity was having a "thick skin". While this mental toughness may apply to professional practice, perhaps, a research study is a different context altogether in the minds of participants; associations of

intellectual credibility, science, honesty, and even integrity, make this a unique space in their working life. Indeed, it is an act of considerable altruism to give time, and genuine creative effort to a study like this, and there in a vulnerability?

The emphasis on the ethics of not showing artwork is none the more present than in the final points of this section. As a research thesis for the development of an argument it would be considerably easier to show the work, and let the artworks speak for themselves. It would also undermine everything that has been stated in this section so far.

Table 60: Extract from fieldwork notes

...today a judge completed an assessment when I noticed they had left an artwork to one side.

"What about this one," I said.

The reply, "I thought that was an error?"

"That's an artwork," I say.

"Aha," they add, with recognition and surprise, "Okay, give me a minute."

It takes longer than a minute, and they end up placing this in the highest creativity category...

...After all the judging is completed, I raise what they thought the word was, and they answer correctly. That was the first time this happened, up to this point many judges don't get it, I tell them, and they nod their head, but with a hint of mild surprise.

...I asked about another artwork, and what the word was: they don't know and give various incorrect guesses.

I say "It's the word XXXXX."

"...Okay" they say, but appear no clearer

"Can you see how they did XXXXX," I add.

An aha moment hits them, "Wow!" they are amazed and laugh, "That's clever."

Table 60 describes a judge's typical interaction for certain artworks, with two exceptions. One of the judges got it by themselves; they rated the artwork as highly creative as possible. Another judge agonised over the rating, in the end placing it in the mid-range rather than low, like most of their peers. A mid-range mark perhaps reflected a tension in rating this work; something clever was going on, but what exactly that was remained elusive within the time frame of judging. A mid-range score in that context may not truly be reflective of a moderately good creative output, but of a degree of confusion about the artwork, and possibly goodwill (they would be inclined to give it 1 or 2, but something didn't quite fit with that score, especially relative to the other works; they opt for 3 or 4).

Both those examples are reflective of lower scoring work with a wide deviation in marks. Although looking directly at standard deviation is not part of CAT protocol, the quantitative approach of the CAT allows distinction between works that are seen by all judges as low (1,1,1,1,1,1,1,1), and others that, though rated low, may mask high levels of creativity (1,1,1,1,6,1,1,1). It is an interesting consideration for the future, an important insight, and extends our understanding of the limitations of the CAT. But, the challenge here is not of that nature, it is whether it is ethical to present such a discussion, given the previous debate above. Clearly, no artwork is shown to support this case, and all of the content in the conversations that could identify the artworks have been removed. Most readers will be none the wiser; the exception is the participants in this study. Reading this content, they may be able to deduce the work. That leaves two dilemmas: one, they may be

incorrect and attribute such insights to themselves that are not the case; two, they may be correct, in which case they can infer a rating: a rating that they may not want to know about.

There are valuable insights, but the examples above, and the explanations pose an ethical problem that cannot be resolved other than through debate. Ethically this complies with the terms of consent, and that no artwork would be shown, but it has risks.

8.12 Limitations

The successful recruitment of participants to studies like these is down to the goodwill and altruism of those taking part. To this extent, it is likely those who do participate in research, are not representative of the general population. As a broad observation, the recruitment of participants, whether judges, experts, intermediates or novices, was around a quarter of all those contacted, leaving up to 73 per cent of those contacted unwilling to participate.

Clearly, the reasons for not participating in any research study are numerous though not always complex; simply, lack of time and other commitments is a frequent reply. It was clear, however, that for some of the novices contacted the issue was the research focus itself: creativity, and associations they had with artistic talent. It must be noted that very few of those contacted who do not take part reply back at any length, but from those who do, some insights are possible, and shed light on the additional challenges present in recruitment to design creativity studies like this one.

For example, one of the reasons for not taking part by a novice participant was that they were: "...the least artistic person you could imagine, I would struggle to draw a stick man

to give you some idea of where my skills lie...I suspect I'd be more hindrance than help." Rather than a hindrance for a study like this, quite the opposite, this person would have made an ideal novice participant.

In this regard, the recruitment of participants are prone to a degree of unavoidable bias: towards participants who are confident in their creative abilities, or confident in being challenged, or quite willing to "make a fool of themselves". As such it is questionable to what extent "uncreative" people even entertain the idea of taking part in a study on creativity? This in itself may place a relative cap on the "lower" levels of creativity within this study. For ethical reasons, it had to be made clear to all potential participants that these were studies about design creativity. That may excite some novices to take part, and for the majority, be a reason to avoid.

8.13 Future Studies

In relation to future studies, one key consideration is how to increase participant recruitment, especially from design professional, in order to run studies with a larger sample size. For the most part, in study 5, the aspect of greatest challenge for design professional was taking part in the psychometric tests. This required them to find at least an hour and a half and in some cases two hours of time to participate (and this was on top of the time they had taken to undertake the creativity tasks). For future studies, it would be advisable to use only one ability as an independent variable, in this case, either visualisation, or written comprehension, but not both together.

The further point is that depending on the nature of a future study, it is debatable how valuable it is to test for general intelligence. Certainly, a body of evidence exists to question the correlation between IQ and creativity when it is measured by domain-specific

means. The quality of research design in Study 5 was enhanced through measuring general intelligence, but given the time, cost and possibility of putting off some participants, for larger studies the value of this enhanced research design needs to be viewed pragmatically.

A further consideration for future research is, in spite of guidance by Kaufman et al, who dedicated a complete chapter of their book *Essentials of Creativity Assessment* to the CAT, as did Hennessey, Amabile, and Mueller for the *Encyclopaedia of Creativity*, there appears to remain substantial variation across research studies in relation to CAT method. As highlighted in this thesis, within the CAT literature, some researchers have created instructions that directly ask judges to discount technical execution from their creativity rating. However, some other researchers ask judges to rate creativity alongside technical execution and aesthetic appeal; some only do this the first time they undertake a new CAT task, while some do not distinguish between these. Such examples suggest the necessity for international standards of consistency and transparency. There is a definite need for further nuanced specification, supported by much more empirical research on CAT procedure, and standardisation of guidelines in the use and reporting of the CAT.

Aside from these questions of CAT standards and protocol, a practical point to consider is that the technical challenges to preparing artwork for the CAT as a barrier to future acceptance and growth. There is a need for software to digitise the administration of the CAT, particularly in a graphic design context, but it could equally apply to any domain where viewing a creative output on screen is required. As may be present in the details of this thesis, the CAT has always been seen as labour intensive, and it is even more so in the case of a graphic design CAT. A DigitalCAT could give the means to evolve the next generation of creativity assessment practices.

8.14 Conclusion

The primary focus of this thesis has been to test the relationship between domain competencies and creative output through empirical study. Several authors reflect views on the nature of this relationship that are broad and diverse. Whether these are relatively simple positive or negative linear correlations, or a more complex form of non-linear relationships, the impact of these perspectives can underpin theoretical, educational and economic strategies for individual, group and sector approaches to creative development.

Built upon the research of Cattell, Horn & Carroll's (CHC) theory of human intelligence, and the Consensual Assessment Technique (CAT) in relation to graphic design creativity, this was the first time both these important models and methods had been applied to graphic design. Specifically, this culminated in the question: *To what extent are abilities, like visualisation and written comprehension, correlated with type and image graphic design creativity, when discounting for verbal and visual IQ?*

Where previous related studies had found correlations, in this study, no significant correlations occurred when general intelligence was accounted. As one of the underlying assumptions regarding the relationship between competencies (be they defined as skills, abilities or knowledge) and creativity was that there was a relationship, the results of this research suggest such underlying assumptions may not be correct when tested against reliable measures of creative output.

If no significant relationship exists then, this could have a direct impact on creativity training and educational practice. For example, what the results suggest is that while there is not a correlation within a group as a whole, there were individual cases. For three participants, scoring highly on the type graphic design creativity task was matched with

scoring highly on Marks VVIQ; conversely, for two high creativity scoring participants, this was the complete opposite with low VVIQ scores. Such a pattern was repeated for low creativity scores and those participants with average creativity scores. What such results highlight is that fastidious views on what abilities are important to nurture can be a theoretical stance not fully borne out by empirical study, but based on isolated cases that can shape sector-wide and institutional perceptions.

Clearly, one study is not sufficient to fully test and evaluate the relationship between domain competency and creative output within graphic design. It is in this respect that this thesis has contributed a number of findings that make the replication of future studies possible and relatively less difficult to undertake.

Firstly, prior to this study, the choice of tasks to measure graphic design creativity was not obvious, and findings highlighted that some tasks did not translate well from design education context into experimental research. Therefore, there was little precedent established for a task to measure graphic design creativity using the CAT reliably. This thesis has provided two new tasks that design researchers, and creativity researchers, can use to measure graphic design creativity: described here as the Image Task and the Type Task.

Secondly, while, at the outset, the CAT was supposed to work for graphic design as it had in other design domains, this was not the case in practice. A graphic design CAT appeared more sensitive to research method than had been assumed, and several caveats to methods and protocol were identified. Of particular importance was an instruction to judges to discount technical execution from their creativity assessment. The implications of this finding, however, go beyond graphic design creativity assessment, and

may help to explain the low levels of consensus in other design and creativity studies. It also raises the issue of international standards for the CAT that would be of benefit to all scholars engaged in the science of creativity.

Thirdly, in relation to identifying and operationalizing domain competencies, this thesis has been the first study to map the conceptual connections between O*NET abilities and other graphic design competency models. Specifically, it explored how frameworks that underpinned O*NET (such as CHC Theory, and F-JAS) could be used to prioritise, and operationalize, graphic design competencies for experimental research. The practicalities of this approach appear to be useful. In this thesis, the abilities of Visualisation (Vz), and Written Comprehension (Reading Comprehension (RC) in CHC theory) appeared important to BTEC, NASAD and O*NET's graphic design competency models as a whole. Through the identification of shared terminology, this allowed existing research literature on F-JAS to suggest operational measures, or aided the search for alternative measures.

This thesis makes three distinct contributions:

- identifies suitable tasks to measure graphic design creativity;
- proposes creativity assessment protocols that are congruent to the needs of design domains;
- integrates specialist graphic design competency models with CHC theory.

Through this knowledge, future design creativity researchers will be able to extend and replicate the findings described above, and continue to build a research-informed, evidence-based approach to design education in its many forms.

References

- Amabile, T. M. (1979). Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology*, 37, 221-233.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology*, 43, 997-1013.
- Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualization. *Journal of Personality and Social Psychology*, 45, 357-376.
- Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Westview Pr.
- Amabile, T. M., Goldfarb, P., & Brackfield, S. C. (1990). Social influences on creativity: Evaluation, coaction, and surveillance. *Creativity Research Journal*, 3, 6-21.
- Amabile, T. M., Hennessey, B. A., & Grossman, B. S. (1986). Social influences on creativity: The effects of contracted-for reward. *Journal of personality and social psychology*, 50, 14-23.
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5), 502-508.
- Avitia, M. J., & Kaufman, J. C. (2014). Beyond g and c: The relationship of rated creativity to long-term storage and retrieval (Glr). *Psychology of Aesthetics, Creativity, and the Arts*, 8(3), 293-302.
- Baer, J. (1991). Generality of creativity across performance domains. *Creativity Research Journal*, 4(1), 23-39.
- Baer, J. (1993). *Creativity and divergent thinking: A task-specific approach*. Lawrence Erlbaum Associates, Inc.
- Baer, J. (1996). Does artistic creativity decline during elementary school? *Psychological reports*, 78, 927-930.
- Baer, J. (1998). The case for domain specificity of creativity. *Creativity Research Journal*, 11, 173-177.
- Baer, J. (2012). Domain specificity and the limits of creativity theory. *The Journal of Creative Behavior*, 46, 16-29.
- Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the Consensual Assessment Technique to Nonparallel Creative Products. *Creativity Research Journal*, 16, 113-117.
- Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: The Amusement Park Theoretical (APT) model of creativity. *Roeper Review*, 27, 158-163.
- Baer, J., & McKool, S. S. (2009). Assessing creativity using the Consensual Assessment Technique. In *Handbook of research on assessment technologies, methods, and applications in higher education* (pp. 1-13).
- Barbot, B., Besançon, M., & Lubart, T. (2016). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences*.
- Batey, M. & Furnham, A. (2006). Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs*, 132, 355-429.
- Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and Individual Differences*, 20, 532-535.
- Beghetto, R. A., & Kaufman, J. C. (2007). Toward a broader conception of creativity: A case for "mini-c" creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 1(2), 73-79.
- Bendassolli, P. F., Wood Jr, T., Kirschbaum, C., & Cunha, M. P. (2009). Creative Industries: definition, limits and possibilities. *Revista de Administração de Empresas*, 49(1), 10-18.
- Bennett, A. (2015). The Fundamental Role of Seeing in Design Thinking, Proceedings of Design Thinking in Design Education, Claremont, California.
- Berland, E. (2012). Creativity and education: Why it matters. Retrieved <http://www.wimages.adobe.com/www.adobe.com/content/dam/Adobe/en/education/pdfs/adobe-creativity-education-findings.pdf?trackingid=KFBUZ>
- Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Applied Cognitive Psychology*, 20(2), 239-263.
- Blazhenkova, O., & Kozhevnikov, M. (2010). Visual-object ability: A new dimension of non-verbal intelligence. *Cognition*, 117(3), 276-301.

- Boden, M. A. (1994). What Is Creativity? In M.A.Boden (Ed.), *Dimensions of Creativity* (pp. 75-118). Cambridge, MA: The MIT Press.
- Boyatzis, R. E. (2008). Competencies in the 21st century. *Journal of Management Development*, 27, 5-12.
- Bronson, P., Merryman, A. (2010, October 7). The creativity Crisis. *Newsweek*. Retrieved (Aug 4, 2016) from <http://europe.newsweek.com/creativity-crisis-74665>
- Brown, R. T. (1989). Creativity: What are we to measure? In J.A.Glover & R. R. R. C. R. Ronning (Eds.), *Handbook of Creativity* (pp. 3-32). New York: Plenum Press.
- Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny reading test: Manual for scoring and interpretation, forms G & H*. Riverside Publishing Company.
- Burns, R. B. (1994). Surveying the cognitive terrain. *Educational Researcher*, 23(2), 35-37.
- BTEC. (2010). BTEC Level 3 Nationals specification in Art and Design: Issue 1. Edexcel, Pearson, UK.
- BTEC. (2013). BTEC Higher Nationals from 2010: What are BTEC HNs? Retrieved (June, 15, 2013) from <http://www.edexcel.com/quals/highernationals10/about/Pages/default.aspx>
- Byrne, C. (2011). *Task selection and the consensual assessment technique: Using collage tasks in creativity research*. University of Central Lancashire.
- Cable, V. and Pickles, E. (2010, September 5). A recovering economy requires local remedies. *Financial Times*.
- Casakin, H., & Kreidler, S. (2011). The cognitive profile of creativity in design. *Thinking Skills and Creativity*, 6(3), 159-168.
- Caughron, J. J., & Mumford, M. D. (2012). The Fleishman Job Analysis Survey. *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations*, 231.
- CCSkills. (2009a). National Occupational Standard for Design. Creative and Cultural Skills.
- CCSkills. (2009b). Create visual design: CCSDES10. Creative and Cultural Skills.
- Chermayeff, I., & Geismar, T. H. (2006). *Watching Words Move*. Chronicle Books.
- Christiaans, H. (1992). *Creativity in design: the role of domain knowledge in designing*. TU Delft, Delft University of Technology.
- Christiaans, H. & Venselaar, K. (2005). Creativity in design engineering and the role of knowledge: Modelling the expert. *International Journal of Technology and Design Education*, 15, 217-236.
- Chronicle Books. (2009). Design and Typography for Fall 2009. San Francisco, Chronicle Books. 12-6-2012.
- CIHE (2010). *The Fuse: Igniting High Growth for Creative, Digital and Information Technology Industries in the UK*.
- Climer, N. (Speaker). (2016, Feb 16). The life Scientific [Audio podcast]. Retrieved from <http://www.bbc.co.uk/radio4>
- Collado-Ruiz, D. & Ostad-Ahmad-Ghorabi, H. (2010a). Influence of environmental information on creativity. *Design Studies*, 31, 479-498.
- Collado-Ruiz, D. & Ostad-Ahmad-Ghorabi, H. (2010b). Influence of Environmental Information on Expert-perceived Creativity of Ideas. *Design Creativity 2010*, 71.
- Collins, H. M. (1985). Changing Order: Replication and Induction in Scientific Knowledge. In (London: Sage.
- Collins, H. M. (2010). *Tacit and explicit knowledge*. Chicago: University of Chicago Press.
- Colman Getty (2010). *Press Release: CIHE Task Force urges far-reaching changes to ensure UK is a leader in the creative, digital and information technology industries* London: Colman Getty Consultancy.
- Conti, R., Coon, H., & Amabile, T. M. (1996). Evidence to support the componential model of creativity: Secondary analyses of three studies. *Creativity Research Journal*, 9, 385-389.
- Costanza, D. P., Fleishman, E. A., & Marshall-Mies, J. (1999). Knowledges. In N.G.Peterson, M. D. Mumford, W. C. Borman, P. R. Jeanneret, & E. A. Fleishman (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 71-90). American Psychological Association.
- Cowdroy, R. & de Graaff, E. (2005). Assessing highly-creative ability. *Assessment & Evaluation in Higher Education*, 30.

- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, 18, 391-404.
- Cropley, D. & Cropley, A. (2010). Recognizing and fostering creativity in technological design education. *International Journal of Technology and Design Education*, 20, 345-358.
- Cross, N. (1997). Descriptive models of creative design: application to an example. *Design Studies*, 18, 427-440.
- Cross, N. (2001). Strategic knowledge exercised by outstanding designers. *Strategic knowledge and concept formation III*, 17-30.
- Cross, N. (2006). *Designerly ways of knowing*. Springer London.
- Crossick, G. (2006). Knowledge transfer without widgets: the challenge of the creative economy. *London: Goldsmiths, University of London*.
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. New York.
- Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R.J.Sternber (Ed.), *Handbook of Creativity* (pp. 313-335). Cambridge University Press.
- Dai, D. Y., Swanson, J. A., & Cheng, H. (2011). State of Research on Giftedness and Gifted Education: A Survey of Empirical Studies Published During 1998-2010. *Gifted Child Quarterly*, 55, 126.
- Diakidoy, I. A. N. & Spanoudis, G. (2002). Domain specificity in creativity testing: A comparison of performance on a general divergent-thinking test and a parallel, content-specific test. *The Journal of Creative Behavior*, 36, 41-61.
- Daily Mail Reporter (2011, August 11). Your imagination is not playing tricks on you children really are becoming less creative, study shows. *Daily Mail*. Retrieved from <http://www.dailymail.co.uk/>
- DCMS (2001a). *Creative Industries mapping documents 2001*, London: Department for Culture, Media and Sport.
- DCMS (2001b). *Culture and creativity: The next ten years*, London: Department for Culture, Media and Sport.
- DCMS (2016a). *Creative Industries: Focus on Employment*, London: Department for Culture, Media and Sport.
- DCMS (2016b). *Creative Industries Economic Estimates*, London: Department for Culture, Media and Sport.
- De Bono, E. *Serious creativity: using the power of lateral thinking to create new ideas*. 1992. New York: HarperColins.
- Design Skills Advisory Panel (2007). *Higher Skills for Higher Value* Design Council and Creative & Cultural Skills.
- Dineen, R. & Niu, W. (2008). The effectiveness of western creative teaching methods in China: An action research project. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 42.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational testing service*.
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., & Matthews, K. A. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, 52(5), 187-200.
- Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge Univ Press.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford Books/MIT Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: Bradford Books/MIT Press.
- Feist, G. J. (2004). the evolved fluid specificity of human creative talent. In R.J.Sternber, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: from potential to realization* (pp. 57-82). APA.

- Film Skills Group. (2003). Developing UK film talent: A comprehensive skills survey of the UK film industry. London: Skillset/Film Council.
- Fleishman, E. A., Costanza, D. P., & Marshall-Mies, J. (1999). Abilities (pp. 175-195). In Peterson, Norman G. (Ed); Mumford, Michael D. (Ed); Borman, Walter C. (Ed); Jeanneret, P. Richard (Ed); Fleishman, Edwin A. (Ed), (1999). An occupational information system for the 21st century: The development of O*NET. , Washington, DC, US: American Psychological Association.
- Fleishman, E. A., & Reilly, M. E. (1992). *Handbook of human abilities: Definitions, measurements, and job task requirements*. Consulting Psychologists Press.
- Florida, F. (2002). The rise of the creative class: and how it's transforming work, leisure, community and everyday life. *New York: Basic*.
- Frensch, P. A., & Sternberg, R. J. (1989). Expertise and intelligent thinking: When is it worse to know better? In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 157–188). Hillsdale, NJ: Erlbaum
- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. White Plains, NY: Longman.
- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: BasicBooks.
- Gardner, H. (2006). *Multiple intelligences: New horizons*. Basic Books.
- Gilhooly, K. J. and Green, A. J. K. (1989). Learning problem-solving skills. In: Colley, Ann M. and Beech, John R. eds. *Acquisition and Performance of Cognitive Skills*. Chichester, UK: Wiley, pp. 88–111.
- Gluck, J., Ernst, R., & Unger, F. (2002). How creatives define creativity: Definitions reflect different types of creativity. *Creativity Research Journal*, 14, 55-67.
- Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). *An introduction to systematic reviews*. Sage.
- Green, A. J. K. and Gilhooly, K. J. (1992). Empirical advances in expertise research. In: Keane, M. T. and Gilhooly, K. J. eds. *Advances in the Psychology of Thinking*, Volume 1. Hemel Hempstead, UK: Harvester Wheatsheaf, pp. 45–70.
- Guay, R. (1977). Purdue Spatial Visualization Test-Visualization of Views. *Purdue Research Foundation, West Lafayette, IN*.
- Guile, D. (2010). Questioning the simplistic link between qualifications and labour market entry: new forms of expertise and learning in the creative and cultural sector. *LLAKES Research Paper*, (2).
- Han, K. S. (2003). Domain-specificity of creativity in young children: How quantitative and qualitative data support it. *The Journal of Creative Behavior*, 37, 117-142.
- Han, K. S. & Marvin, C. (2002). Multiple creativities? Investigating domain-specificity of creativity in young children. *Gifted Child Quarterly*, 46, 98.
- Harvey, R. J. & Wilson, M. A. (2010). Discriminant validity concerns with the O* NET holistic rating scales. In.
- Haylock, D. (1997). Recognizing mathematical creativity in school children. *International Reviews on Mathematical Education*, 29, 68-74.
- Hennessey, B. A. (1994). The consensual assessment technique: An examination of the relationship between ratings of product and process creativity. *Creativity Research Journal*, 7, 193-208.
- Hennessey, B. A., Amabile, T. M., & Martinage, M. (1989). Immunizing children against the negative effects of reward. *Contemporary Educational Psychology*, 14, 212-227.
- Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (2011). Consensual assessment. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (2nd Ed., Vol. 1) (pp. 253-260). San Diego, CA: Academic Press.
- HM Treasury (2005). *The Economic and Fiscal Strategy Report and the Financial Statement and Budget Report* (Rep. No. HC 372). London: The Stationary Office (HM Treasury Publication).

- HM Treasury (2015). *Fixing the foundation: Creating a more prosperous nation* (Rep. No. Cm 9098). London: The Stationary Office (HM Treasury Publication).
- Horn, J. (1998). A basis for research on age differences in cognitive abilities. In J. J. McArdle, & R.W.Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57–92). Mahwah, NJ: Lawrence Erlbaum.
- Howard, T. J., Culley, S., & Dekoninck, E. A. (2011). Reuse of ideas and concepts for creative stimuli in engineering design. *Journal of Engineering Design*, 22, 565-581.
- Isaksen, S. G., Murdock, M. C., Isaksen, S. G., Murdock, M. C., Firestien, R. L., & Treffinger, D. J. (1993). The emergence of a discipline: Issues and approaches to the study of creativity. *Understanding and recognizing creativity: The emergence of a discipline*, 13-47.
- Isaksen, S. G., & Treffinger, D. J. (2004). Celebrating 50 years of reflective practice: Versions of creative problem solving. *Journal of Creative Behavior*, 38, 75-101.). Celebrating 50 years of reflective practice: Versions of creative problem solving. *Journal of Creative Behavior*, 38, 75-101.
- IET (2015). Skills and demand in industry 2015 survey, the Institution for Engineering and Technology, London.
- Ivcevic, Z. (2009). Creativity Map: Toward the Next Generation of Theories of Creativity. *Psychology of Aesthetics Creativity and the Arts*, 3, 17-21.
- Jackson, D. (2016). Modelling graduate skill transfer from university to the workplace. *Journal of Education and Work*, 29, 199-231.
- Jeffries, K. K. (2007a). Diagnosing the creativity of designers: individual feedback within mass higher education. *Design Studies*, 28, 485-497.
- Jeffries, K. K. (2007b). Identifying domain relevant skills. In *10th European Conference on Creativity and Innovation*.
- Jeffries, K. K. (2009). *Skills for Creativity in Games Design (Part 1): Academic Conceptions of Creativity in Games Design*. Brighton: Higher Education Academy: Art, Design and Media subject centre.
- Jeffries, K. K. (2011). Skills for creativity in games design. *Design Studies*, 32, 60-85.
- Jeffries, K. K. (2012). Amabile's Consensual Assessment Technique: Why has it not been used more in design creativity research? *Proceedings of the 2nd International Conference on Design Creativity (ICDC2012)*, 1, 211-220
- Jeffries, K. K. (2015a). A CAT with caveats: Is the Consensual Assessment Technique a reliable measure of graphic design creativity? *Proceedings of the 3rd International Conference on Design Creativity (ICDC2015)*, 12th-14th January 2015, Institute of Science, Bangalore, India.
- Jeffries, K. K. (2015b). A CAT with caveats: Is the Consensual Assessment Technique a reliable measure of graphic design creativity? *The International Journal of Design Creativity and Innovation*.
- Jensen, A. R. (2004). Obituary—John Bissell Carroll. *Intelligence*, 32(1), 1–5.
- Julmi, C., & Scherm, E. (2015). The Domain-Specificity of Creativity: Insights from New Phenomenology. *Creativity Research Journal*, 27(2), 151-159.
- Kaufman, J. C. (2015). Why Creativity Isn't in IQ Tests, Why it Matters, and Why it Won't Change Anytime Soon Probably. *Journal of Intelligence*, 3(3), 59-72.
- Kaufman, J. C. & Baer, J. (2002). Could Steven Spielberg Manage the Yankees?-Creative Thinking in Different Domains. *The Korean Journal of Thinking & Problem Solving*, 12, 5-14.
- Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative?. *Creativity Research Journal*, 24(1), 83-91.
- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20, 171-178.
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of general psychology*, 13(1), 1.
- Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken, NJ US: John Wiley & Sons Inc.

- Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, 47, 635–650.
- Kelly, W. F., Clark, A. C., & Branoff, T. J. (2013). Spatial test correlation in an introductory graphic communications course. *Proceedings of the Web of Graphics, 68th Mid-year conference of the Engineering Design Graphics Division*, 20th-22th October 2013, Worcester Polytechnic Institute, Worcester, MA.
- Kokotovich, V. (2008). Problem analysis and thinking tools: an empirical study of non-hierarchical mind mapping. *Design Studies*, 29, 49-69.
- Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualisers: A new characterization of visual cognitive style. *Memory & cognition*, 33(4), 710-726.
- Kozhevnikov, M., Kozhevnikov, M., Yu, C. J., & Blazhenkova, O. (2013). Creativity, visualization abilities, and visual cognitive style. *British journal of educational psychology*, 83(2), 196-209.
- Kim, K. H. (2011). The creativity crisis: The decrease in creative thinking scores on the Torrance Tests of Creative Thinking. *Creativity Research Journal*, 23(4), 285-295.
- Larrue, F., Sauzéon, H., Foloppe, D., Wallet, G., Cazalets, J. R., Gross, C., & N'Kaoua, B. (2013, September). Assessing the impact of automatic vs. controlled rotations on spatial transfer with a joystick and a walking interface in VR. In *IFIP Conference on Human-Computer Interaction* (pp. 1-18). Springer Berlin Heidelberg.
- Lau, K. W. (2011). The difficulties of assessing design students' creativity: a critical review on various approaches for design education. *Journal of Design Research*, 9, 203-219.
- Lawson, B. (1994). *Design in mind*. Architectural Press.
- Lee, J. H., Gu, N., & Ostwald, M. J. (2015). Creativity and parametric design? Comparing designer's cognitive approaches with assessed levels of creativity. *International Journal of Design Creativity and Innovation*, 3(2), 78-94.
- Levy, C. & Hopkins, L. (2010). *Shaping up for innovation: are we delivering the right skills for the 2020 knowledge economy?* National Centre for Vocational Education Research (NCVER).
- Lipsett, A. (2008, June 24). Video games degrees: 95% fail to hit skills target. *Guardian*.
- Lubart, T. & Guignard, J. H. (2004). The generality-specificity of creativity: A multivariate approach. *Creativity: From potential to realization*, 43-56.
- Macdonald, I. W. (2006). Mindset and Skillset; the persistence of division in media education. *Journal of Media Practice*, 7, 135-142.
- Mackie, R. K. (2005). Multiple intelligences and graphic design ability in five North Carolina community colleges. (Unpublished doctoral dissertation). North Carolina State University, Raleigh.
- MacKinnon, D. W. (1962). The nature and nurture of creative talent. *American psychologist*, 17(7), 484.
- Marks, D. F. (1972). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64, 17-24.
- Masterson, J., & Hayes, M. (2004). UK data from 197 undergraduates for the Nelson-Denny Reading Test. *Journal of Research in Reading*, 27(1), 30-35.
- Masterson, J., & Hayes, M. (2007). Development and data for UK versions of an author and title recognition test for adults. *Journal of Research in Reading*, 30(2), 212-219.
- Mayer, R. E. (1999). Fifty Years of Creativity Research. In R.J.Sternberg (Ed.), *Handbook of creativity* (pp. 449-460). Cambridge: Cambridge University Press.
- Mazilescu, C. A., Mitroi, M. M. P., Draghici, A., Mihartescu, A. A., & Pop, M. C. (2010). Representation of teaching competency to students and teachers. *Annals of DAAAM & Proceedings*, 19-21.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford.
- McGrew, K. (2004). Cattell-Horn-Carroll CHC (Gf-Gc) Theory: Past, present & Future. <http://www.iapsych.com/CHCPP/CHCPP.html>. Obtido, 17(03), 2006.

- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1-10.
- McKee, S. (2012). Reading Comprehension, What We Know: A Review of Research 1995 to 2011. *Language Testing in Asia*, 2(1), 1-14.
- Molloy College. (2014). Admissions & Aid: Frequently Asked Questions about the Nelson-Denny Exam. Retrieved from <http://www.molloy.edu/>
- Mosoloane, R. A. (2013). *Textbook diagrams illustrating phases of the moon: Grade 10 learners' interpretation in relation to spatial ability* (Doctoral dissertation).
- Nagai, Y. & Gero, J. (2012). Special issue on Design Creativity.
<http://www.jaist.ac.jp/ks/labs/nagai/DesignCreativityW/> . 7-3-2012.
- NASAD. (31-1-2012). NASAD Handbook 2011-12. Virginia, National Association of Schools of Art and Design.
- Nathan, M., Pratt, A., & Rincon-Azner, A. (2015). Creative economy employment in the EU and UK: A comparative analysis. *London: Nesta*.
- Nesta. (2015, Dec 15). Retrieved from <http://www.nesta.org.uk/news/uk-accounts-fifth-eus-creative-industry-jobs>
- Nickerson, R. S. (1999). Enhancing Creativity. In R.J.Sternberg (Ed.), *Handbook of creativity* (pp. 392-430). Cambridge, UK: Cambridge Univ Press.
- O*NET Online. (2016, Aug 2). Retrieved from <http://www.onetcenter.org/overview.html>
- Oxman, R. (2002). The thinking eye: Visual re-cognition in design emergence. *Design Studies*, 23, 135–164.
- Pektas, S. T. (2010). Effects of cognitive styles on 2D drafting and design performance in digital media. *International Journal of Technology and Design Education*, 20, 63-76.
- Peters, T. J. (2003). *Re-imagine! Business Excellence in a Disruptive Age*. London: Dorling Kindersley.
- Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y. et al. (2001). Understanding work using the Occupational Information Network (O* NET): Implications for practice and research. *Personnel Psychology*, 54, 451-492.
- Peeverly, S. T., & Sumowski, J. F. (2012). What variables predict quality of text notes and are text notes related to performance on different types of tests?. *Applied Cognitive Psychology*, 26(1), 104-117.
- Peeverly, S. T., Garner, J. K., & Vekaria, P. C. (2014). Both handwriting speed and selective attention are important to lecture note-taking. *Reading and Writing*, 27(1), 1-30.
- Peeverly, S. T., Vekaria, P. C., Reddington, L. A., Sumowski, J. F., Johnson, K. R., & Ramsay, C. M. (2013). The Relationship of Handwriting Speed, Working Memory, Language Comprehension and Outlines to Lecture Note-taking and Test-taking among College Students. *Applied Cognitive Psychology*, 27(1), 115-126.
- Plucker, J. A. (1998). Beware of simple conclusions: The case for content generality of creativity. *Creativity Research Journal*, 11, 179-182.
- Plucker, J. A. (1999). Reanalyses of student responses to creativity checklists: Evidence of content generality. *The Journal of Creative Behavior*, 33, 126-137.
- Plucker, J. A. (2004). Generalization of creativity across domains: Examination of the method effect hypothesis. *Journal of Creative Behavior*, 38, 1–12
- Plucker, J. A., & Beghetto, R. A. (2004). Why Creativity Is Domain General, Why It Looks Domain Specific, and Why the Distinction Does Not Matter. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity From potential to realization* (pp. 153-167).
- Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. *Handbook of creativity*, 35-61.
- Polanyi, M. (1958). *Personal knowledge: towards a post-critical epistemology*. Chicago, IL: University of Chicago Press.

- Pontillo, T. M. (2010). *Spatial ability and handedness as potential predictors of space teleoperation performance* (Doctoral dissertation, Massachusetts Institute of Technology).
- Puccio, G. J. (2002). Foursight: the breakthrough thinking profile-Presenter's guide and technical manual. *Evanston, IL: THinc Communications*.
- Rashidi, N., & Khosravi, N. (2010). Assessing the Role of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension of Iranian EFL Learners. *Journal of Pan-Pacific Association of Applied Linguistics*, 14(1), 81-108.
- Reiter-Palmon, R., Illies, M. Y., Cross, L. K., Buboltz, C. B., & Nimps, T. (2009). Creativity and domain specificity: The effect of task type on multiple indexes of creative problem-solving. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 73.
- Ridderstråle, J., & Nordström, K. (2004). *Karaoke capitalism: Management for mankind*. Pearson Education.
- Robson, C. (1993). *Real world research: A resource for social scientists and practitioner-researchers*. Oxford, UK: Blackwell Publishers.
- Roy, R. (1993). Case studies of creativity in innovative product development. *Design Studies*, 14(4), 423-443.
- Runco, M. A. (1999). Creativity need not be social. *Social creativity*, 1, 237-264.
- Runco, M. A. (2004). Everyone Has Creative Potential. In R.J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realisation* (pp. 21-30). Washington, DC: American Psychological Association.
- Valgeirsdottir, D., Onarheim, B., & Gabrielsen, G. (2015). Product creativity assessment of innovations: considering the creative process. *International Journal of Design Creativity and Innovation*, 3, 95-106.
- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3rd ed.) (p. 99-144). New York: Guilford.
- Scott, G., Leritz, L. E., & Mumford, M. D. (2004). The effectiveness of creativity training: A quantitative review. *Creativity Research Journal*, 16, 361-388.
- Shaw, T. H., Matthews, G., Warm, J. S., Finomore, V. S., Silverman, L., & Costa, P. T. (2010). Individual differences in vigilance: Personality, ability and states of stress. *Journal of Research in Personality*, 44(3), 297-308.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science*, 191, 952-954.
- Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review*, 27(4), 599-606.
- Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is Creativity Domain-Specific? Latent Class Models of Creative Accomplishments and Creative Self-Descriptions. *Psychology of Aesthetics Creativity and the Arts*, 3, 139-148.
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I. et al. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68.
- Skillset (2003). *National occupational standards: Directors*. London: The Sector Skill Council for the Audio Visual Industry.
- Skillset. (2004a). Govt spotlight on media employers productivity & skills needs. London: Skillset (The Sector Skills Council for the Audio Visual Industry)
- Skillset (2004b). *Job profiles for the audio visual industries: 2004* London: The Sector Skill Council for the Audio Visual Industry.
- Skillset (2005). *Job profiles for the audio visual industries: 2005* London: The Sector Skill Council for the Audio Visual Industry.
- Skillset (2008). *Undergraduate Course Accreditation Guidelines for Computer Games*.
- Skillset (2009). *Creative Media Workforce Survey 2008* Skillset.
- Skillset (2011). *Sector skills assessment for the creative industries of the UK*.

- Sorby, S. A. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education*, 31(3), 459-480.
- Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, 26, 20-29.
- Stacey, M. K., Eckert, C. M., & Wiley, J. (2002). Expertise and creativity in knitwear design. *International Journal of New Product Development and Innovation Management*, 4, 49-64.
- Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal*, 18, 87-98.
- Sternberg, R. J. & Lubart, T. I. (1995). *Defying the crowd*. The Free Press.
- Thornton, T. (2012). Tacit Knowledge. Symposium conducted on ethnographic research in health studies at UCLan, Preston, UK, 17-5-2012.
- Torrance, E. P. (1966). The Torrance tests of creative thinking-norms-technical manual research edition e verbal tests, forms A and B e figural tests, forms A and B. Princeton, NJ, Personnel Press.
- Torrance, E. P. (1995). *Why fly?: A philosophy of creativity*. Norwood, NJ: Ablex Publishing Corporation.
- Torrance, E. P. (1998). The Torrance tests of creative thinking norms: technical manual figural (streamlined) forms A & B. Bensenville, IL, Scholastic Testing Service.
- Universities UK (2010). *Creating Prosperity: the role of higher education in driving the UK's creative economy* London: Universities UK.
- Valgeirsdottir, D., Onarheim, B., & Gabrielsen, G. (2015). Product creativity assessment of innovations: Considering the creative process. *International Journal of Design Creativity and Innovation*, 3(2), 95-106.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*.
- Verstijnen, I. M., Van Leeuwen, C., Goldschmidt, G., Hamel, R., & Hennessey, J. M. (1998). Sketching and creative discovery. *Design Studies*, 19, 519-546.
- Wall, T., Coley, S., Davies, F., Dubber, A., Harte, D., and Long, P. (2006). Leeds MeCCSA conference. *Three-D*, 13-15.
- Waller, A. A., & Lourenco, S. (2010). Proceedings from 40th ASEE/IEEE frontiers in education conference: Special session—Frontiers of spatial cognition: Moving beyond rotating cubes. Washington, DC: ASEE/IEEE (American Society of Engineering Education/Institute of Electrical and Electronics Engineers).
- Ward, T. B., Patterson, M. J., & Sifonis, C. M. (2004). The role of specificity and abstraction in creative idea generation. *Creativity Research Journal*, 16, 1-9.
- Weisberg, R. W. (2006). Modes of Expertise in Creative Thinking. In KA Ericsson, N Charness, PJ Feltovich, & RR Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (Cambridge, UK: Cambridge University Press).
- Wilson, M. A. (2007). A history of job analysis. In L. L. Koppes (Ed.), *Historical perspectives in industrial and organizational psychology* (pp. 219-241). Mahwah, NJ: Erlbaum.
- Wojtczuk, A. (2014). *Creative product assessment in design: Influence of judges' backgrounds and levels of experience in design* (Doctoral dissertation, Aix-Marseille).

Appendix A: Example of Consent Forms for Study 4



Artwork Consent Form

Title of the project: Creativity in Graphic Design

Main investigator and contact details: Karl K. Jeffries, 01772 895 185, or at kjeffries@uclan.ac.uk

1. I have read the Information Sheet which is attached to this form, I understand what my role will be in this research, and agree to take part.
2. I understand that I am free to withdraw from the research up to the point of judges rating the artworks: this will begin on Jan 2nd, 2013.
3. I have been informed that the confidentiality of the information I provide will be safeguarded.

Name of participant (print).....Signed.....Date.....

Appendix B: Randomisation of Type Task Artwork for Study 4

Randomisation of Type artwork: 30 Type artworks were coded from 10-39 and randomised into 16 sets (coded from Set A to Set P). Each set was then randomly allocated to each judge.

Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H	Set I	Set J	Set K	Set L	Set M	Set N	Set O	Set P
23	33	31	14	12	34	24	14	23	25	30	37	24	24	18	22
29	19	34	19	13	22	37	34	38	29	28	21	30	29	19	12
31	38	15	21	36	16	26	33	26	18	38	35	31	21	39	35
32	36	14	38	31	23	17	29	13	19	19	31	14	12	35	19
12	16	21	18	24	28	10	30	19	22	33	19	27	31	12	38
37	24	36	20	18	31	18	17	17	12	29	12	34	32	33	13
13	10	18	22	16	18	32	22	33	31	37	33	26	36	34	21
20	37	13	12	15	15	22	18	37	10	31	38	28	17	16	30
25	17	28	31	29	33	31	36	14	15	20	24	25	14	17	34
30	20	35	15	21	29	19	26	31	38	36	16	39	39	26	27
27	28	24	29	10	14	30	11	12	21	35	22	13	25	29	17
39	25	37	23	34	25	11	35	21	39	16	20	15	33	32	23
33	34	38	28	35	36	25	28	20	28	32	30	35	11	36	31
14	22	22	32	14	26	34	38	30	20	11	13	18	27	11	36
24	15	32	35	37	35	13	31	28	34	21	28	36	20	10	37
19	23	20	24	19	32	12	24	24	37	39	39	11	30	27	26
34	32	12	34	26	27	28	15	32	14	17	26	37	18	37	14
35	39	30	36	32	12	36	16	15	32	14	27	33	37	20	28
17	35	29	30	11	10	38	19	11	23	13	14	20	13	22	39
11	13	23	37	38	19	29	13	29	17	22	18	16	38	13	29
15	29	39	27	23	13	20	12	27	13	18	11	29	22	21	32
22	18	10	13	17	21	35	25	39	27	25	34	21	15	38	33
38	31	27	25	30	39	14	10	35	36	10	15	17	28	24	11
21	27	25	39	28	30	21	20	22	24	12	17	22	23	28	16
10	11	19	26	25	20	23	32	18	26	23	36	12	19	25	15
16	26	16	17	33	37	33	23	16	30	27	23	38	26	14	20
36	14	11	10	22	17	39	27	36	33	24	29	32	16	31	25
18	30	26	11	20	11	16	21	25	11	34	32	23	35	30	24
28	21	33	33	27	38	15	39	10	16	26	25	10	10	15	18
26	12	17	16	39	24	27	37	34	35	15	10	19	34	23	10

Appendix C: Randomisation of Image Task Artwork for Study 4

Randomisation of Image artwork: 30 Image artworks were coded from 41-70 and randomised into 16 sets (coded from Set A to Set P). Each set was then randomly allocated to each judge.

Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H	Set I	Set J	Set K	Set L	Set M	Set N	Set O	Set P
68	54	59	67	59	45	48	52	60	57	64	46	47	53	47	64
58	49	49	42	63	51	53	67	49	42	68	57	66	45	58	48
60	67	68	62	45	64	54	46	67	41	52	43	45	69	48	68
45	50	63	68	52	54	62	63	58	60	50	42	60	48	61	54
43	52	45	57	49	48	46	43	70	66	60	50	58	62	65	70
46	70	48	50	56	61	51	47	63	44	61	44	67	56	59	63
50	69	66	47	48	53	55	61	51	59	53	70	56	44	49	62
62	64	55	66	68	66	63	54	57	45	41	62	55	55	64	47
51	43	42	59	53	60	59	42	47	52	51	61	53	49	45	56
48	44	43	43	47	68	70	68	68	43	57	47	61	63	70	51
67	51	56	49	55	70	47	64	61	62	54	41	52	52	43	60
64	59	70	46	58	41	58	65	45	48	58	65	70	67	55	65
42	58	60	53	50	44	41	66	52	51	48	59	65	68	52	50
70	60	65	58	62	55	43	48	50	56	62	63	68	51	57	45
41	66	57	45	46	50	42	59	59	63	47	56	49	50	41	46
63	46	69	61	65	52	60	41	44	49	67	48	41	61	62	49
56	41	46	54	61	69	67	53	48	54	49	66	50	65	46	52
59	53	53	52	41	47	68	55	64	58	66	64	63	41	66	67
54	42	64	55	44	62	49	50	43	53	55	67	69	43	63	44
65	61	58	44	42	59	64	44	69	50	69	58	64	58	42	59
44	63	50	65	70	42	56	60	42	61	59	60	44	59	50	61
53	45	62	41	43	65	65	70	66	65	65	69	46	54	67	58
49	48	51	51	64	43	66	58	56	47	43	45	54	66	53	42
69	47	67	69	60	46	52	62	55	70	70	53	51	70	54	69
47	57	44	63	66	49	69	57	46	46	46	52	59	64	69	41
66	65	47	48	57	56	57	49	65	69	63	68	57	46	56	57
57	68	54	70	51	67	45	45	53	68	56	51	62	47	60	55
52	62	41	64	54	58	50	69	41	67	45	49	43	42	68	43
61	55	52	56	67	63	61	51	62	64	44	55	48	57	51	53
55	56	61	60	69	57	44	56	54	55	42	54	42	60	44	66

Appendix D: Image Task Technical Instructions for Study 5



Information on Technical Support for the Image task

As this image task is designed to offer a creative challenge across the full range of graphic design expertise, it is important that the tools available to create artwork provide the greatest freedom of expression. To this end, an industry standard used by professional graphic designers and students is Adobe Illustrator.

For participants in this study who are unfamiliar with Adobe Illustrator, technical support will be available. This technical support is to enable each participant to create an artwork that reflects his or her creative vision. It is vital however, that this technical support only deals with the practical aspects of transferring a participants design to Adobe Illustrator, rather than offering advice or suggestions on the quality of the design.

To safe guard against such influence, technical support will be offered remotely via email. For example, a participant may send two photographs/images via email to kjeffries@uclan.ac.uk. These images will then be transferred to Adobe Illustrator, and this digital illustration will be sent back to the participant. Any amendments required by a participant can be suggested, and the cycle will continue until the participant is satisfied, or until the time limit on the task has expired.

The time required to create this artwork is expected to take no more than one hour. However, thinking time and playing with ideas may increase for some participants.

This task is not timed in a conventional sense but, for practical purposes, a limit of one week from receiving the task has been set. How productive or efficient graphic design ideas can be create is not part of this study, and one week will allow all participants the chance to think, reflect and play with their ideas, or engage with the technical support option described above, prior to submitting a single artwork via email.

A time limit of one week will allow participant's ideas to incubate, which is often an opportunity lacking in many experimental creativity research studies. The reason for this is the desire to control for the influences of other individual's ideas on the participants own work. To mitigate this influence, we ask that you do not discuss your ideas with others prior to completion within the week. In this way, artwork can only be attributed to your own efforts, which is exactly what is required for this study.

Unless otherwise stated by a participant, the size of the image of both images will be cropped to 10cm by 10cm, and will be cropped relative to the centre of the image. Previous participants have found https://www.google.co.uk/advanced_image_search useful; please select "Large" for the image size if you use Google Advance Image Search.

Appendix E: Example of Information Sheet for Study 5

Project title: Skills for Creativity in Graphic Design



The Open
University

Hi,

I'm writing to ask if you would be willing to participate in a research study on skills for creativity in graphic design. Below are key questions and answers to help you decide whether to participate. Thank you for taking the time to consider this.

What is this research about?

This study is to understand whether particular skills underpin creativity in graphic design. Specifically, it will explore the relationship between visualization (the ability to form mental images) and reading comprehension. A previous study found that reading ability appears to be more important for creativity in graphic design than visualisation, but some of the methods used in that study may have influenced this conclusion. We would like to revisit this with more appropriate measures of visualisation and graphic design creativity.

Why have I been invited to participate?

Thirty participants are required whose current experience in graphic design will range from novice to expert, and who live within a 30 mile radius of Preston.

What will happen if I take part?

Your participation will take place over three weeks, at a time that is convenient to yourself, and at a location of your choice. The only person you will have contact with will be myself. In week one and two you will be given two graphic design tasks that require creativity with either images you have selected, or a word you have selected. You can do this in your own time and convenience, and we will be in contact by email and phone if you would like any technical support on creating these artworks. In the third week we will meet at a location of your choice, and I will provide a number of visualisation and reading comprehension tasks for you to complete within a certain time, but no longer than an hour and fifteen minutes in total.

The artwork you create will be seen by eight full-time professional graphic designers, who will be asked to rate each example for graphic design creativity. The visualisation and reading measures will be scored by myself following the procedures required.

How much of my time will this require?

During the three weeks of the study, we expect that the tasks will take one hour a week to create each week. We anticipate the total time of your participation will be between three to three and a half hours of your time.

Will the information I provide be kept confidential?

Yes, all information you provide will be treated in confidence. The only person who will know you have taken part in the study will be myself. Your name will not be used in relation to any of the artwork or results. No artwork will be reproduced in any publications without asking for your written consent first.

What will happen if I want to stop taking part?

If you wish to withdraw, I will respect your decision immediately. You will be able to withdraw up to the point of judges rating the artworks: this will begin on the 1st October, 2014. The artworks and results up to the period of withdrawal may be used, if you are happy for this to occur. Otherwise you may request the works and results to be destroyed and no further use made of them.

What will happen to the results of this study?

The findings of the research will be written up as feedback for you, for policy makers and for other organisations interested in our work. The findings will be published as part of a 2016 PhD thesis, as journal and conference papers, and they may also be used for teaching and research training.

Appendix F: Example of consent form for Study 5



Participant Consent Form

Title of the project: Skills for Creativity in Graphic Design

Main investigator and contact details: Karl K Jeffries, 01772 895 185, or at kjeffries@uclan.ac.uk

1. I have read the Information Sheet which is attached to this form, I understand what my role will be in this research, and agree to take part.
2. I understand that I am free to withdraw from the research up to the point of judges rating the artworks: this will begin on 1st Oct 2014.
3. I have been informed that the confidentiality of the information I provide will be safeguarded.

Name of participant (print).....Signed.....Date.....